

## SEARCH REQUEST FORM

Requestor's Name: \_\_\_\_\_ Serial Number: \_\_\_\_\_  
Date: \_\_\_\_\_ Phone: \_\_\_\_\_ Art Unit: \_\_\_\_\_

**Search Topic:**

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

## STAFF USE ONLY

Date completed: 10-07-02  
Searcher: Beverly C 4994  
Terminal time: 25  
Elapsed time: \_\_\_\_\_  
CPU time: \_\_\_\_\_  
Total time: 25  
Number of Searches: \_\_\_\_\_  
Number of Databases: 1

## Search Site

\_\_\_\_\_ STIC  
\_\_\_\_\_ CM-1  
\_\_\_\_\_ Pre-S

## Type of Search

\_\_\_\_\_ N.A. Sequence  
\_\_\_\_\_ A.A. Sequence  
\_\_\_\_\_ Structure  
\_\_\_\_\_ Bibliographic

## Vendors

\_\_\_\_\_ IG Suite  
\_\_\_\_\_ STN  
\_\_\_\_\_ Dialog  
\_\_\_\_\_ APS  
\_\_\_\_\_ Geninfo  
\_\_\_\_\_ SDC  
\_\_\_\_\_ DARC/Questel  
\_\_\_\_\_ Other CGN



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 02:19:23 ; Search time 1114.8 seconds  
(without alignments)  
11996.956 Million cell updates/sec

Title: US-09-736-968A-1  
Sequence: 1 gagcgagagaccatgctgc.....aagcaaaaaaaaaaaaaa 6372

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenBank:  
1: gb\_da:\*  
2: gb\_hlg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: gb\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hlg\_hum:\*  
31: em\_hlg\_inv:\*  
32: em\_hlg\_other:\*  
33: em\_hlg\_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result Query Score Match Length DB ID Description

1	6372	100.0	6372	6	AX173175	AX173175 Sequence
2	4882.8	76.6	4886	9	AB037816	AB037816 Homo sapi
3	2797.8	43.9	2825	9	BC008335	BC008335 Homo sapi
4	2140	33.6	2149	6	AX173233	AX173233 Sequence
5	2132.8	33.5	6828	6	AX172874	AX172874 Sequence
6	1729.4	27.1	7215	6	AX174569	AX174569 Sequence
7	1468.2	23.0	4311	6	AB051558	AB051558 Homo sapi
8	1446.6	22.7	4144	6	AX172880	AX172880 Sequence
9	1446.6	22.7	4144	6	AX172932	AX172932 Sequence
10	1446.6	22.7	4200	6	AX172980	AX172980 Sequence
11	1357.6	21.3	1900	10	BC021414	BC021414 Mus muscu
12	1225.4	19.2	3327	6	AK055905	AK055905 Homo sapi
13	1207	18.9	3233	6	AX172978	AX172978 Sequence
14	1167.8	18.3	4577	6	AK024436	AK024436 Homo sapi
15	1031.8	16.2	4026	6	AX174571	AX174571 Sequence
16	1031.8	16.2	4027	6	AX174664	AX174664 Sequence
17	1007	15.8	2452	6	AK055493	AK055493 Homo sapi
18	865.8	13.6	2014	6	AX174662	AX174662 Sequence
19	815.6	12.8	2904	9	BC019102	BC019102 Homo sapi
20	597.8	9.4	2607	9	AK055401	AK055401 Homo sapi
21	408.4	6.4	73515	2	AC017375	AC017375 Drosophi
22	407.4	6.4	166626	3	AC008318	AC008318 Drosophi
23	407.4	6.4	301639	3	AE003590	AE003590 Drosophi
24	400.8	6.3	1242	9	AF194407	AF194407 Homo sapi
25	372.2	5.1	434	6	AX329840	AX329840 Sequence
26	322.4	5.1	3899	6	AX058220	AX058220 Sequence
27	308.4	4.8	2998	10	BC009134	BC009134 Mus muscu
28	274.6	4.3	3227	10	RNTRG	X68101 R. norvegicu
29	255.2	4.0	5862	9	AB028981	AB028981 Homo sapi
30	255.2	4.0	7506	6	AX255048	AX255048 Sequence
31	254.8	4.0	123321	9	AC011472	AC011472 Homo sapi
32	249	3.9	37107	9	AC009000	AC009000 Homo sapi
33	247.8	3.9	1669	9	HSMB02936	HSMB02936 Homo sapi
34	238.2	3.7	321	6	AX173254	AX173254 Sequence
35	238.2	3.7	3023	9	AK022412	AK022412 Homo sapi
36	216.6	3.4	281	6	AX173240	AX173240 Sequence
37	213.8	3.4	1605	6	AX174666	AX174666 Sequence
38	203.6	3.2	2413	9	BC015018	BC015018 Homo sapi
39	199.8	3.1	2036	9	AK001253	AK001253 Homo sapi
40	198.4	3.1	2299	9	AK054649	AK054649 Homo sapi
41	198.2	3.1	2422	9	AK000227	AK000227 Homo sapi
42	195.2	3.1	382	6	AX341959	AX341959 Sequence
43	194.8	3.1	2610	6	AX058192	AX058192 Sequence
44	194.8	3.1	3152	9	AK056684	AK056684 Homo sapi
45	194.8	3.1	3472	6	AX173120	AX173120 Sequence

## ALIGNMENTS

RESULT	1					
LOCUS	AX173175	6372 bp	DNA	linear	PAT 03-JUN-2001	
DEFINITION	Sequence 1 from Patent WO0142295.					
ACCESSION	AX173175					
VERSION	AX173175.1	GI:14598036				
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
TITLE	1 (bases 1 to 6372)					
JOURNAL	Clasp-7 transmembrane protein					
	Patent: WO 0142295-A 1 14-JUN-2001;					
	Arbor Vita Corporation (US)					
FEATURES	Location/Qualifiers					
source	1..6372					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	/note="full length human CLASP-7 cDNA"					
	13..6156					
	/note="human CLASP-7"					

CDS

		/codon_start=1 /protein_id="GACA3731.1" /db_xref="GI:14598037" /translation="MAASERRAFNAHKINRTVAENVRKQVSRERSGSPHSSRRSSSLG VPLREVEAPLEDEVDLSPRPDAEPGLRDVLVFPDDELILLQRECKRTTGPJPKD EKLDQVRAAVEMVIEDWYIVHRYQVLTADYDIDREOKLAPROVFEODAG DERSGPEDNSRSGSPEDTPRSSASIFPLRLNADSLPILSEAEAEEDNR NETLRORHPALLTLYPADDEDAVRERPREHNGORLYVCHLSLKEIEIEP IFGLIAYDREKKIISENFPDLSMNGSLIARCTAPALSTARSLASVATPSP DIRITILEVLOOGDISSECEPYMLKEDTKAKNEKLEKRLAEQDCTLGKRM PFAVTVAHLAVISSAGLDSDSDSBERNPATDRKRGDORASGGDACSFGFK PATLVTFNFKOEAEKLSDEDLKFLADPAEVPSSLRLRQVPTAOLKIDISAPENPH FCLSPDLHLIKYPDPGRPKTEILFPAEYVAAPTSYRNLYVPHSLNNSRQGS VRNLAVVOYMTGEDPSQALPIFGKSSCEPTEARETVVYHNKSPFEYEFKLLP ACVTENHLLFTFYHNSQORPGTALETTPVGFMTIILQHRKRTCPCLPVSQOP PSYSLTPDALPQMRWDGKHGVSVELGASVPEOPYLDKPEITVPIVEEGAP FRKDVTLSGNEOEELRASLALRLASPRIVAFSHYHLDKRVIRITPPIISOIV NLGCAFEAMAHVSYIHRSLAODARCKPOLAAVYHARLPETPEPSLPDGPVY TVQAKTILARSGRPAALILARSKSISSNPDILAVAGSVDDERSILAKLHEELAL QWVSSAVSRBALIQHAWFEPQLMVKSMALHLILQRLDTPKRLRPGREFLDITALV GSVGLVITRVHNDVLEHNLASLAEFLSDLSLDVDRGFVSLVAHKQVATRIQS SPNPAALITLMETTRILCSHEHYVTLNPPCPLSPAPSPSVSTSSQSTFSSOA PDPKVTSMFELSGPFRQHFAGLILTELALAEPAEGAPSLHKKATSAVSLICGH DTPDRVAETATKARVAELYLILISTARDTLPRIHDAEAGRGORSRLASMLSDTEGSG DIAGTINPSVAMALAGPRLA PGSRASISQSPPTASAGALSBSRTILAEVLYLK NTEPALQRMATDITLPLQGRLLDLTLCLAEYIGKKAFFKINSLIFKKSLLDKAR LEBAIILGTICARQDMVRSRERSPPGPNPENVRKRSVTHMKQISDVHDTKDMHEA LEBNLATEASLVVLDLLEIIVQVLMSEARSLGAVLKVLVLSGASQSLFLOHG LATGRALVSKPELLEPDEDLACADLCRLRHGSGRISTIRTHASASLYLIMRQNE IGHNFAVRKMQWMSLSLVGTQNFSEELRSLRTILTYAEDMLGDMSTFAEVO DLNFNLMLITDVTYKMEHQDEPMILDMYRIARQYSGPDLRLIWMONMGKAEI GNHAEACQVHAALVAEYLLALEDOHRLPVCVSFROUISVLEESISDILSPD EEGRCSKHPTELGLVGLLEQANGYFTMGSLYVAANVEYKNLPILEARDYKRIIAV HGKIQEAFKIMHOSGWERFEGTFRVGTGAHFDLDEQEVTLKEPSTLAEIASH RLKEFYTEREGDDVETIKDSYVPVSKSLDSQXAYIQTIVVEYEPITLKRKRETVFD RNLGRTLEFCTPEPTPDGAHAGELPEQHKRKLPLSDHAKFPIYKTRIRKCHEEYVLI PVEVALIEDMQKTRIELAFATODPDAKMLQYLOQSVPTGQGEVLEAQAFLAEP EDPKLFRHNKRLKCFKDFCKCEDALRRKKALIGDPQKEYHRELBNYCRLEALQIP LLQRLPOLAMAPPPGLRNSLNASFPKADL"		BASE COUNT 1273 a 2050 c 1849 g 1200 t ORIGIN	
Query Match		100.0%; Score 6372; DB 6; Length 6372;			
Best Local Similarity 100.0%; Pred. No. 0;					
Matches 6372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
Qy	1	gacgcgagagaccatgctgcgctccgagcgccgcttcgcgcacaaagatcaacaagag	60		
Db	1	GACGCGAGGACCATGGCTCTCCGAGCGCCGCTTCGGCGCAAGATCAACAGGACG	60		
Qy	61	gtgcgcgacaggtgcgaaagcaagtgtcccggaacgcagctgctcccccactccagc	120		
Db	61	GTGGCGGACGAGAGTGGCGAAGCGAGGTGTCCGGGAGACGAGTGGCTCCCCCACTCCAGC	120		
Qy	121	aagcgctgcagcagctccctctgggggtcccaactgactgaagtgtcgagcccttgacct	180		
Db	121	AAGCGCTGCGAGAGCTCCCTCGGGGTCCCACTGACTGAAGTGTGCGAGCCCTCGACTTT	180		
Qy	181	gaggaatgactctgagccgcccacacagatgtcagcccgccccccaaggaagcttgta	240		
Db	181	GAGGATGACTTCTGAGCGCGGCGCACAGATGTGAGCCCGCGCCCTCAAGGACCTGGTA	240		
Qy	241	gaattcccaagctgacttgagagctgctgctgagcccccgggaatgctcggaaccaagag	300		
Db	241	GAATTCCCAAGCTGATGATCTTGAGACTGTGCTGACGCCCGGGGAAATCCCGACACGGAG	300		
Qy	301	cccgagatcccccaagatgaaaaactgatacccaagtgagggcggggtggagatgta	360		
Db	301	CCCGGATCCCCAAGATGAAGAACTGATGCCAAGGTGAAGGGCGGGGTGGAGATGTAA	360		
Qy	361	attgaggaatggatctgtccacagaagatcagtaactgagatgagatacacagccc	420		
Db	361	ATTGAGAACTGGATCTTGTCCACAGAAGATCACTAGTGTGACATACACCCC	420		
Qy	421	gtcaccacagacacacagcgggagcgacagaaagggcctcccccgaagtccttgagcag	480		
Db	421	GTCACCAACAGACACAGCGGAGCGACGACAGAAAGGCGCTCCCGCAGAGCTTTGTGACAG	480		
Qy	481	gatgtcttgagaaagcggagagtcggccctcggaagacccgaatgaccccgagctgctg	540		
Db	481	GATGCTTTGGAAGACGAGAGAGTCCGGCCCTGAGACTCGAATACCTCCCGCGTGGCTCG	540		
Qy	541	ggctcccccgaagaacacccctcgaaagcagtggtgcctctagcalettcgaaactgagaa	600		
Db	541	GGCTCCCCGGAAGACACCCCTCGAAGCAGTGCTCTGTGCACTTTCGACTGAGAGAAC	600		
Qy	601	ctggagagctgactatgactgctccctctctgctaagcgggcgccccagaagaatgtagac	660		
Db	601	CTGGAGCTGACTATTGCTCCTCTCTCTGTAAGCGGGCGCCCCAGAAATGTGGAC	660		
Qy	661	cggcgcaatgaaaccccttcgacggcagcaacggcccccggccctgtcaacctctacccg	720		
Db	661	CGGCGCAATGAACCCCTTCGACGCGCAGCACCGGCCCCCGCCCTGTCAACCCCTACCCG	720		
Qy	721	gcacctgacgagatgtagccgctggaacgctgtgagccgcccagaagccaccccgagac	780		
Db	721	GCACTTGACGAGATGAAGCCGTGAACCGCTGAGCCGCCAAGCCACCCCGCAGAC	780		
Qy	781	tttgacacaaagatcttgatcaagtgctcgtcgctcaagttcgagatltgaattgagccc	840		
Db	781	TTTGACACAAAGATCTTGATCAAGTGTCTGCTCGCTCAAGTTGCAATTTGAGCCC	840		
Qy	841	atctcttgagacttgagcctctgtatgctgtgcggaagaaagaaagatctcgagaaacttc	900		
Db	841	ATCTTTGGAATCTTGCTCTGTATGATGTGCGGAGAAAGAAAGATCTCGGAACTTC	900		
Qy	901	tacttcgacctgaactcggagactccaatgaaaggggctgttcgggtctaatgagacacccct	960		
Db	901	TACTTCGACTGAACCTGGACTCCATGAAGGGCTGCTTGGGCTCATGGACCCACCT	960		
Qy	961	gcaatctcacaccctggcgccgctctgcacatctctctgtgacccaaccccaactgacac	1020		
Db	961	GCAATCTCACACCCTGGCGCCGCTCTGCACATCTCTCTGTGACCTTACCTTACCTGACATC	1020		
Qy	1021	ttcctgtcatcaagtgtgagaaagtgctctcaagcaaaaggagacatcagtgagtgtgtgag	1080		
Db	1021	TTCTGTGATCAATGATGGAAGAGTGTCTTACAGAAAGGACATCAGTGTGCTGTGAG	1080		
Qy	1081	ccttacatgtgtgtgaaagaggttgacacagcccaagaacaaagaaagcttagaagaagctg	1140		
Db	1081	CCTTACATGTGTGAAAGAGGTGACACAGCCAAAGAAAGAACTTAGAAGAGCTG	1140		
Qy	1141	cgcttcgcgcgacgacgtcttcacccgccttgagccgctacccgcatgacctcgacctg	1200		
Db	1141	CGCTTCGCGGCGGACGAGTCTTCTGCACCCGCTGGCCGCTTACGCAATGCCCTTCGCGCTG	1200		
Qy	1201	acggcgtgtgaacttggccaacatcgtgtgacagcgctgggacgtgtgacccggaaacttgac	1260		
Db	1201	ACGGCGTGACCTTGGCCAAATCTGTGAGCAGCGCTGGGACGTGACCGGACTTGTGAC	1260		
Qy	1261	tcggagggcgagcgccggccagccttgacagacacgcgcgctcgaggggccccagagcgag	1320		
Db	1261	TCGGAGGGCGAGCGCCCGCAGCCTTGACAGACCGCGCGCTTGAGGGGCCCAAGAGCCGG	1320		
Qy	1321	gcgagtagtgggagcagcgcctgcaagctctcgtgcttcgctccagccaagctaaactgtc	1380		
Db	1321	GCGAGTAGTGGGAGCAGCGCCTTGACAGCTTCTGCGCTTCCGTCACACCGCTAACTGTGC	1380		
Qy	1381	acaaactctttaagcagaggggctgagcagactgaagagagaaacctcttaagttacggt	1440		
Db	1381	ACAAACTCTTTAAGCAGAGGCTGAGCAGACTGAGTGAAGAGAACTCTTCAAGTTCTCG	1440		
Qy	1441	gtctgacatgaagcggcccgctgctccctgtgtcgcgcgagactaagtcgtgtgactgcccagctc	1500		
Db	1441	GCTGACATGAAGCGCCCGTGTCTCTGCTGCGGCACTAGTCTCTGTGACTGCCAGCTC	1500		



QY 1501 aagatcgaacattctctcgcgtctctgaaaaatcccaactctgcctctccctcgaagtgcct 1560  
 |||||  
 Db 1501 AAGATCGACATTTCTCCGGCTCTGAAATGCCACTTGTGCTCTCCCTGAGCTGCTT 1560  
 QY 1561 catatcaagccctaccgggaaccccaaggcccgcccaagaagatatctcgtgagttcccc 1620  
 |||||  
 Db 1561 CATATCAACCCCTACCCGGACCCCGAGGGCCCGCCACCAGAGATTTCTGAGTTCCCC 1620  
 QY 1621 gcccggaagttctatgccccccataccagcttacaggaactcgtctagttaccggcac 1680  
 |||||  
 Db 1621 GCCCGGAGAGTTCTATGCCCCCATACACACTACAGAACCTGCTTACTGTGACCCGGAC 1680  
 QY 1681 agcctcaacttcaagcagcccgcaaggctctcgtcgcgaaactcgtcgtcgaagtgcagttac 1740  
 |||||  
 Db 1681 AGCCTCAACTTCACGACGGCGGACAGGCTCCGTCGCAACCTGTGCTGAGTGCAGTAC 1740  
 QY 1741 atgacagcgagagagaccccaagcgtctcgcgtgtcactctttgggaagtccaagctgcagat 1800  
 |||||  
 Db 1741 ATGACAGCGGAGAGACCCCGACAGGCTCTGCGGTATCTTTGGCAAGTCCAGCTGCAGT 1800  
 QY 1801 gaattaccgcgagagccttcaacacggtgtgtctacataaagaatcccccgaagttctac 1860  
 |||||  
 Db 1801 GAATTTACCCGGGAGGCTTCAACACGGTGTGCTACACATTAACAAGTCCCGGAGTTTAC 1860  
 QY 1861 gaagagttcaaggtgcatcttccaagcctcgtgaaagagaaccatacactcgtcttcaac 1920  
 |||||  
 Db 1861 GAGGAGTTCAAGTTCATCTTCCAGCTTCGCTGACAGAGAACCATACCTGCTGTTTCAC 1920  
 QY 1921 tcttaccatgttcagctcgtcagaccccgccggcgagcactgcctctggagagacacccgttgagctt 1980  
 |||||  
 Db 1921 TTCTACCATGTCTACGTGACGACCCCGCGGCGGACTGCTTGGAGACACCCGTTGGGCTTT 1980  
 QY 1981 acttgatcccaactcgtcgtcagacagggcgctcgtgaagagccgcctctcgtctcccaatg 2040  
 |||||  
 Db 1981 ACTTGATCCCACTGCTGAGAGACAGGGCGCTGAGAGACCGGCCCTTGTCTCCACAGTG 2040  
 QY 2041 tctcgtgagcaagcgccgcgcgcagctatcttcgtgtctcaaccccgatgtgtgccttccgggc 2100  
 |||||  
 Db 2041 TCTGTGAGACAGCGCGCGGCCACAGCTATTCGCTGCTCACACCCGATGTGGCTTCGGGGC 2100  
 QY 2101 atgcgcttggtgagagtgtaacaagaagcgtgttcaagtgtgagagctcaacgcgtgtccctc 2160  
 |||||  
 Db 2101 ATGCGCTGGGTGACGGTCAACAAGGGCGTGTTCAGTGTGAGAGCTCACAGCCGTGTCTCT 2160  
 QY 2161 gtgcaccccccaagjacccctactcgtgacaaatcttcaaccctgtgtgcagcttccctggaagag 2220  
 |||||  
 Db 2161 GTGCAACCCCGAGGACCCCTACCTGACAAATCTTCAACCCGTGTGACAGTCCCTGGAGAG 2220  
 QY 2221 ggaagccttcccaatcccggtcgaaggaacactgtgtcgtgaagcgaggagcaagtgtagcagaag 2280  
 |||||  
 Db 2221 GGAGCCTTCCCATTCGCGGTCAAGGACACTGTGCTAGGAGGAGCAACGTGAGGACAGGAG 2280  
 QY 2281 ctgcggggcagctcttgcaagaactgcgtgcggcagcagcccgaaaccccttctgtgcctctcc 2340  
 |||||  
 Db 2281 CTGCGGGGCCAGTCTTGTGAGACACTGCGCCTGTGCGACGCCGGAACCCCTTGTGCGCTTCTCC 2340  
 QY 2341 caaccaactgctgtaacaagctcgtgtcgtlctgtlcaatcaaggcccccgaatcaatcgaatgagcag 2400  
 |||||  
 Db 2341 CACACAGCTGCTGAGCAAGCTCGTGTGCTGATCATCAGGCGCCCGATCATCATGAGTGCAG 2400  
 QY 2401 atgtgtgaactgtggtccgtgtgaagccttgaagcaatgagcccaatgtatgaagccttgttcaac 2460  
 |||||  
 Db 2401 ATGTGTGAACCTGGGCGGTGAGGCTTTGAAGCAATGAGGCCCATGTATGTAGCCCTTGTTCAC 2460  
 QY 2461 cggagccttgaggagagcccgagatgcgccggtgcactgtcccaagctgtgcctcgtc 2520  
 |||||  
 Db 2461 CGGAGCTGAGAGGACGCCAGATGCGCGGTGCTACTGCCCAAGAGCTGGGTGCTTACGTG 2520  
 QY 2521 caactaagccttgcgtctcctgtgcaactgaagccagcctcccggaatggtggccctccagtg 2580  
 |||||  
 Db 2521 CACTAAGCCTTGTGCGCTTCTGTGCACTGAGCCAGCCTCCCGGATGGGGCCCTTCCAGTG 2580

QY 2581 acagtgcaagctgccaacactgcccgtgtgctgtgtcgcgcccgcaagcctctacactgcgcg 2640  
 |||||  
 Db 2581 ACAATGCAAGCTGTGCAACACAGGCCGCTGTGTGCTGCGCCGCCAAGCCTTACCTGTGGCG 2640  
 QY 2641 cgttccaagagcatcaagcagcaagaacccagactcgcgtcgcgttgccctgtgcctcgtgagat 2700  
 |||||  
 Db 2641 GCTTCCAAGAGCATACGACAGCAACCCCTGACTCGCCCTGTGGCCCTGTGTGTGAGAT 2700  
 QY 2701 gacgaagtcttccgcacatccctgcccagcaagctcgtctcaagagagactcgtcgtcagtg 2760  
 |||||  
 Db 2701 GACGAGGTTTCCGCACTGCGACAGCAAGCTGTTACAGAGAGAGTGGCTGTGAGTGG 2760  
 QY 2761 gtgtgtcagcagcagtgccgtgaacgcgagagagccactccagagacgcctggttctcttccaa 2820  
 |||||  
 Db 2761 GTGTGTGACAGCACTGCGTACAGGAGGCGCATCTCCAGACCCCTGTGTCTTCTCCAG 2820  
 QY 2821 ctcatgtgtgaagatgagcgtcgtcagctcgtcgtcgttgcagagagactagaacaccccg 2880  
 |||||  
 Db 2821 CTCAATGCTGAAGAGTATGGCGCTGCACCTGCTGTGGCCAGGACTAGACACCCCGC 2880  
 QY 2881 aagctgcgcttcccccgaagcgttccctggaagcaatcaatcgtccttgtgtggtcctgtgggc 2940  
 |||||  
 Db 2881 AAGCTGGGCTTCCCGGACAGCTTCTGAGACATCACTGCTGTGTGGCTGTGTGGC 2940  
 QY 2941 ctggaagtcaatcaccggtgttccacaagaagtgtggaagctggtgcagacaccccaagcagc 3000  
 |||||  
 Db 2941 CTGAGAGTATACCCGCTGCTCACAGAGATGTGAGCTGGCCAGACACTCAACGCCAGC 3000  
 QY 2941 ctggaagtcaatcaccggtgttccacaagaagtgtggaagctggtgcagacaccccaagcagc 3060  
 |||||  
 Db 2941 CTGAGAGTATACCCGCTGCTCACAGAGATGTGAGCTGGCCAGACACTCAACGCCAGC 3060  
 QY 3001 ctggaagtcaatcaccggtgttccacaagaagtgtggaagctggtgcagacaccccaagcagc 3060  
 |||||  
 Db 3001 CTGAGAGTATACCCGCTGCTCACAGAGATGTGAGCTGGCCAGACACTCAACGCCAGC 3060  
 QY 3001 ctggaagtcaatcaccggtgttccacaagaagtgtggaagctggtgcagacaccccaagcagc 3120  
 |||||  
 Db 3001 CTGAGAGTATACCCGCTGCTCACAGAGATGTGAGCTGGCCAGACACTCAACGCCAGC 3120  
 QY 3061 gtccgggcccactcaagcaggtgtgccaagcgtgtccagtcgttccctaaatccaagagcc 3180  
 |||||  
 Db 3061 GTCCGGGCCACTCAAGCAGGTGGCCACGCGGCTCCAGTGTCTTAAATCCAGCAGCC 3180  
 QY 3121 ctgctgacccctgacatgaaattcaaccgcacatcctgtgcagccagcaactacgtgtac 3180  
 |||||  
 Db 3121 CTGCTGACCCCTGCGATGGAATTCAACCGCATCTCTGTGACAGCCAGACATACGTATAC 3180  
 QY 3181 ctcaacttccctctgtgccccctgtlcaactcagcactcgcctccctcctgtgtctcc 3240  
 |||||  
 Db 3181 CTCAACCTCCCTGTGTGCCCCCTGTGACCTTCACGCTTCGCCCTTCCCTGTGCTCTCC 3240  
 QY 3241 accaacttcccaagactccaacttctcaagccaaagcccggaaccccaaggtgagcaagatg 3300  
 |||||  
 Db 3241 AOCACCTCCAGAGCTCCACCTTCTCCAGCCAGCCCGGACCCCAAGAGTACAGCATG 3300  
 QY 3301 ttggaactgaatggaacattccggcagagacttccatagcttggtgtcctcgtcgtgaagag 3360  
 |||||  
 Db 3301 TTGCAACTGATGTGACCAATTCCGGCAGGACACTTCTAGCTGGGCTCTCTGTACGGAG 3360  
 QY 3361 ctggaactgagccttgcgaactcgtgaagcgtgaagggcattcctgtgtgcaagaagcactc 3420  
 |||||  
 Db 3361 CTGGAAGCTGAGTGTGACCAATTCCGGCAGGACACTTCTAGCTGGGCTCTCTGTACGGAG 3420  
 QY 3421 agtgtgtgcaagcgtctctatgtgtgcataagcactgaaccccgccctagccgcgaagcact 3480  
 |||||  
 Db 3421 AGTGTGTGCAAGCTCTGATGTGAGCCATGACACTGACCCCGCTACGCGAGGCACT 3480  
 QY 3481 gtgaagcctggtgtgcccgaagctgaactgcagctcgtcttccgalttgcaaggaataacttg 3540  
 |||||  
 Db 3481 GTGAAGGCTGCTGTGTGGCGGAGCTGTGACTGTGCACTGTTTGAATTTGCAAGGAGTACTG 3540  
 QY 3541 ccaagcctgacatgacttgcgtgaagggccaggtcagcggltcaagaactggtcctaagtctc 3600  
 |||||  
 Db 3541 CCACGGCTGACATGACTTGTGTAGGGGCCCAAGGTACAGCGGTCAAACTGCGCTCAATCTT 3600  
 QY 3601 gactaagaacaaagaaagggaaagggagacattgcgggtacacatacaacccctcgtgtggcagt 3660  
 |||||  
 Db 3601 GACTAGACACAGAAAGGAGGAGGAGCAATGTGGGTACATCAACCCCTCTGTGTGGCATG 3660  
 QY 3661 gccattgctgtgtgccccctagccccctgtcctccggcgcaagcatctccaggggccaaca 3720

```
Db 3661 GCCATTGCTGGTGGCCCTTAGCCCTGGCTCCGGGGCCAGCATCTCCCAAGGGCCACCA 3720
Qy 3721 acggtcttcgcgcagagctgtgcccctctctgtctagtacaaacggagactgtgcgtgtc 3780
Db 3721 ACGGCTTTCGGCGAGGCTGTGCCCTCTCTCTCACTCAATCAACCGGACCTTGTGCTGTGT 3780
Qy 3781 gtgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3840
Db 3781 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3840
Qy 3841 aacactcccccaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3900
Db 3841 ACACCTCCCCCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3900
Qy 3901 tacaaggggaaaaagagcttttgaacgcatacaagagcttccatccaataatctctgtat 3960
Db 3901 TACAAGGGGAAAAAGGCTTTGAAAGCATCAACAGCCTCAATTCATAAAAAATCTCTGGAT 3960
Qy 3961 atgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4020
Db 3961 ATGAAGCGCGCGCGCTAGAGGAAGCCATTCTGGTACATCGAGCTCGACAAGAAATGTGT 4020
Qy 4021 cgcgcgaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4080
Db 4021 CGCGGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4080
Qy 4081 gtccacacactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4140
Db 4081 GTCCACACACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4140
Qy 4141 gaggccttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4200
Db 4141 GAGGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4200
Qy 4201 gagatcacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4260
Db 4201 GAGATCATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4260
Qy 4261 ctgtgaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4320
Db 4261 CTGTGAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4320
Qy 4321 ctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4380
Db 4321 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4380
Qy 4381 gaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4440
Db 4381 GAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4440
Qy 4441 atccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4500
Db 4441 ATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4500
Qy 4501 cacaactttgcccgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4560
Db 4501 CACAACCTTTGCCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4560
Qy 4561 acgcgaagacttcaagtaagaagacactgtgcagcttcaactaaaacatctccatactgt 4620
Db 4561 ACGCAGAACTTCAAGTAAAGACACCTGTGCAGCTTCACTCAAAACATCTCACTCAATGCT 4620
Qy 4621 gtaggagagacatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4680
Db 4621 GTAGGAGAGACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4680
Qy 4681 aaactgtcacatgtatctctgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4740
Db 4681 AAACCTGTCAATGTATCTCTGACGACACCGGTGAAGATGAAGGAACACCAAGAGACCTCTAG 4740
Qy 4741 atgtctcatcgacctatgttacagaatgtgtcccggtgttaccaggtgtaccagacctgtcg 4800
Db 4741 ATGTCTCATCGACCTATGTTACAGAATGTGCCCGGGGCTACCAGGGCTCACCGGACCTTGG 4800
Db 4801 ctgtacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4860
Db 4801 CTGTACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4860
Qy 4861 gcccaatgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4920
Db 4861 GCCCAATGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4920
Qy 4921 cagcgccacactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4980
Db 4921 CAGCGCCACACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4980
Qy 4981 gtagtcgcacatcttcgaagacactctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5040
Db 4981 GTAGTCGCACATCTTCGAAGACACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5040
Qy 5041 caattcaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5100
Db 5041 CAATTCACTGACACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5100
Qy 5101 ggcgcgccttctacgagcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5160
Db 5101 GCGCGGCTCTACGAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5160
Qy 5161 caacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5220
Db 5161 CACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5220
Qy 5221 atcatgtcacacagagcttccgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5280
Db 5221 ATCATGTCAACAGAGTTCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5280
Qy 5281 taagcgccccactctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5340
Db 5281 TACGCGCCCCACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5340
Qy 5341 acgaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5400
Db 5341 ACGAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5400
Qy 5401 gaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5460
Db 5401 GACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5460
Qy 5461 aaggtcttaccatccagatcaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5520
Db 5461 AAGGCTTACATCCAGATCAAGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5520
Qy 5521 cgggtgtacctacttgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 5580
Db 5521 CGGGTGAACCTACTTGAACGCGCAACTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5580
Qy 5581 acgcgcgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5640
Db 5581 ACGCGGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5640
Qy 5641 agaacgcgaacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 5700
Db 5641 AGAACGCACACACGCGCTTCCCTCAATCAAGATCGCATCGCTGTGTGTGTGTGTGTGTGTGT 5700
Qy 5701 acggtgtctgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 5760
Db 5701 ACGGTGTCTGAACGCGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5760
Qy 5761 gaccttgcacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 5820
Db 5761 GCCCTTGTGCACCGAGCAGGACCCACAGATGTCTAAGATGTGTGTGTGTGTGTGTGTGTGTGT 5820
Qy 5821 tctgttaggcgcacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 5880
Db 5821 TCTGTAGGCCCCAACCGTGAACCAAGGCTCCCTGTGAGGTGCCCGGCTGTGTGTGTGTGTGTGT 5880
```

QY 5881 atcccggaagaccaccaagctctccgcatcacaaatgcygctgtctcaaggac 5940  
|||||  
Db 5881 ATCCCGGAAGACCCCAAGCTCTCCGGCATCACAAATTTGGGCTCTGCTTCAAGAC 5940  
QY 5941 ttctgcaagaatagtgaagatgctgctgcggaataaagaagccctgattggcggaacag 6000  
|||||  
Db 5941 TTCTGCAAGAAATGTGAGATGCTGCGGAAATAAGAGCCCTTATTTGGGCGGACAG 6000  
QY 6001 aaggagatcaccctgagatgtagcgaactatctccctcctgcggaagctctgagccc 6060  
|||||  
Db 6001 AAGGAGTACCACCTGAGTGAAGGCACTACTGCGCTGCGGAGGCTCTGCAAGCCC 6060  
QY 6061 ctgctatcccaagcgcctgcgcccagctgtagcgaaccaccccccgcgctcagaagatcc 6120  
|||||  
Db 6061 CTGCTATCCACAGCGCTGCGCCAGCTGATGCAACCCACCCCGGCTCAGAACTCC 6120  
QY 6121 ttgacaagagcaagttccggaagaagcaactctgagcccaagaagcaagctgtacc 6180  
|||||  
Db 6121 TTGCAACAGAGCAAGTTTCCGAAAGGCACTCTGAGCCCAAGACCAAGCTGTACC 6180  
QY 6181 tagagaacacaaagaccgagcctcaagctgctgtctgcgagaggagctgcctctgtgac 6240  
|||||  
Db 6181 TAGAGGAACACAGACCCGGGCTCTGCTGTGCTGCGAGGAGTGTGCTGCTGCTCC 6240  
QY 6241 ccactggtctgtgagtgtagcacaactgtagctgagggctgagccctctgtgtccc 6300  
|||||  
Db 6241 CCACGTGGCTGTGAGGTGACACACTGTACTGTGGGCTGGGCTGCTGCTGTGCTCC 6300  
QY 6301 catctgtgtgcatgtgtgtctctccctctttaaatttaaaatggttttaataagcaaaa 6360  
|||||  
Db 6301 CACTGTGTGCACTGATGCTCTCTCTCTTTTAAATTTAAATGTTTATTAAGCAAAA 6360  
QY 6361 aaaaaaaaaa 6372  
|||||  
Db 6361 AAAAAAAAAA 6372

RESULT 2  
LOCUS AB037816 4886 bp mRNA linear PRI 14-MAR-2000  
DEFINITION Homo sapiens mRNA for KIAA1395 protein, partial cds.  
ACCESSION AB037816  
VERSION AB037816.1 GI:7243170  
KEYWORDS  
SOURCE Homo sapiens brain cDNA to mRNA, clone.lib:PbluescriptII SK plus  
clone:hj07927.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS 1 (sites)  
Nagase,T., Kikuno,R., Ishikawa,K.I., Hikosawa,M. and Ohara,O.  
TITLE Prediction of the coding sequences of unidentified human genes.  
XVI. The complete sequences of 150 new cDNA clones from brain which  
code for large proteins in vitro  
JOURNAL DNA Res. 7 (1), 65-73 (2000)  
MEDLINE 20181126  
REFERENCE 2 (bases 1 to 4886)  
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.  
TITLE Direct Submission  
JOURNAL Submitted (31-JAN-2000) Osamu Ohara, Kazusa DNA Research Institute,  
Laboratory of DNA Technology; 1532-3 Yana, Kisarazu, Chiba  
292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp,  
URL:http://www.kazusa.or.jp/huge/, Tel:+81-438-52-3913,  
Fax:+81-438-52-3914)  
FEATURES  
source  
1. .4886  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="hj07927"  
/tissue\_type="brain"  
/clone\_lib="PbluescriptII SK plus"  
1. .4884  
gene

CDS  
/gene="KIAA1395"  
<1..>4884  
/gene="KIAA1395"  
/note="Start codon is not identified."  
/codon\_start=1  
/product="KIAA1395 protein"  
/protein\_id="BA092633.1"  
/db\_xref="GI:7243171"  
/translation="DARTMAASERRAFAHKINRTVAAYEVKOVSRERSGSPHSRRCS  
SSLGVLTEVEVLEDFEDVLSPPDAEPGLADVEFPADDELLDLPRECRTEPG  
IPKDEKLDQVRAVAVYEDMYIVRRQVLSAASPYTDTORERQGLPROVED  
DASGDSRGSGPDSNDSSRGSPEDTPRSSGASISPDNLNADSLPDLRAVED  
VDRNETLRQHRNPALLLTLPRAPDDEAVERCSPREPRHRQATIIKCSITFEI  
ETIPFIFGLALVDREKKITSEFTDLSNDKGLKLRHGHTRPALTLSRATISVT  
YSPDIFLVTKLEKVLQGDIDSECEPVYALVVDKAEVLEKRLAEQFRLG  
RYRMPFAMTVHLNANIVSSAGQLDRSDSEGERPAMTDRRRRGPDRAISGDCSF  
SGRPATLVITNFKEOERLSDLEDFKLDMRRSSSLRLRLRPVDAOLKIDISPA  
ENHPCISPELLIKIPYDPGRPTKEILEEPREYVAPHTSYRLNLYVPSLNEFS  
RQSVYNRLAVRQVYMGDEPSQALPYIRKSSCSPTREAFTPVYHNSSPFEYEFK  
LHLPACTEENHLFTFYHSCOPRGCTALEPVTGTYWPILOHRRLPRGPRCLAVS  
DOPPSYSVLTDPVALPGMKWDGKGVSEVLTAVSVPDPTLDKFTLVHLE  
GAPPRFKDVLISEGNVEQLRASLALRLAEPDLVAFSHVLDKLVLRPILIS  
GQIVNLGRGAFEMAHVSVLVRSLAADARQHPOLAAVYVAFRLGTEPSLDG  
APPVVOAATLARGSPASLYLARSKSISSNPDLAVAPGSYDEVSRIILASKLHE  
ELALQVYSSAVREAILLOHAFEPOLMKSMALLLGORLDPKRLPGRPLDPI  
TALVGSGLVETTRVHKDVELAHLNALSAPFLSDLSLYDGSFVPSIYRAIKQVAT  
RLDSSPAPALLTLRMEFTRLCSHSHVYTLNLECCPLSPSPSPSVSTISOSTP  
SSQAPDPKVTSMELSGPPEQOFLAGLLTELALALEBAEAFLLHKAISVHSL  
LCHGDDPRYAEATVARKAVALELYPLLSIARDPLPHIDPABGGRSILHAKISDST  
EGBGDLAGTINPISVMAIAGGPLAPGRASISOGPTARACGALSSESRLACVL  
WVKLNPBALORWADTLPLPOLGRILDLIYLICLAEAFKKAKFERISISLFEKSD  
MKARLEALILGTIGARQEMVRRRESRPENPNVWRKSYHMKQTSRQVKTDEM  
EHEALVEGNLATEASLVLDLETIVQVWLSRARSVIGAVLYLYLSGNSQSLF  
LQHLGARALVSKFPELLFEDELTCAADCLRLRSGRSISTITNHSASLYLMR  
ONFEIIGHNFARVMQVYMSLSLVGTQONSEHRLRSKTIITVAEEDMGRLDSTFA  
EOYVDLNFNLMLITDPTVKMKEHQEDPEMLIDIMYLRARGQSPDLRLTWLQNMAGX  
HAEIIGNAEAAOCMVNAA"

BASE COUNT 941 a 1605 c 1425 g 915 t  
ORIGIN

Query Match 76.6%; Score 4882.8; DB 9; Length 4886;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4884; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gacgcgagacatgctgctccgagcgcgcgccttcgcgcaagaatcaagagc 60  
|||||  
Db 1 GACGCGAGACCATGGCTCCCTCCGAGCGCGCGCTTCGCGCAAGATCAAGAGACG 60  
QY 61 gtggcgcgagaggtgycggaagaagtgctccgggaagaagcagtgctcccccacatcagc 120  
|||||  
Db 61 GTGGCGGAGAGGTGCGGAGAGGTGTCGGGGAACCAAGTGGCTCCCCCATCTCAGC 120  
QY 121 aagcgcctgcaagcagctccctgggggtcccaactgtagtgaagttgctgagccctgagctt 180  
|||||  
Db 121 AGCGCCTGAGAGAGCTCCCTGGGGGTCCCACTGACTGAAGTTGTCAGACCCTGAGACTTT 180  
QY 181 gagagtgtaactctgaagccgcgcagcagatgctgagcccgaggccctcaagagctgta 240  
|||||  
Db 181 GAGAGTGACTTCTGAGCGCGCACCAATCTGAGCCCGGCGCCCTTGAGGACTTGTA 240  
QY 241 gaattcccaagctgtagtctgtagctgtagcagcccgagggaatgtagcagcaagag 300  
|||||  
Db 241 GAATTCACAGCTGATGACTTGGAGCTGCTGCTGCAAGCCCGGGAATGCCGACACGAG 300  
QY 301 cccgggatccccaagaagatgaatgtagtcccaagctgtaggagcccgagtgagagatgat 360  
|||||  
Db 301 CCGGGATCCCAAGAGAAATGATGATGCCCGAGGAGGCGCGCTGGAATGAT 360  
QY 361 attgagagctgggtcatctgctcaagagagatataagctcccgagtgtgagagataagcccc 420  
|||||  
Db 361 ATTGAGAGCTGGTCAATGTCACAGAAAGTATCACTGAGTGCAGCATACAGCCCC 420

```
Qy 421 gtacacacagacacacagcgagagcgaaagggcctcccccgccaggtctcttgagcag 480
Db 421 gtacacacagacacacagcgagagcgaaagggcctcccccgccaggtctcttgagcag 480
Qy 481 gatctctctgagagcgagaggtccgagcctgagagactcgatatactccggtgtgctcg 540
Db 481 gatctctctgagagcgagaggtccgagcctgagagactcgatatactccggtgtgctcg 540
Qy 541 ggtctcccggaagacacccctcgaaagcagtgctgctctagcatctctcgaccgaggaac 600
Db 541 ggtctcccggaagacacccctcgaaagcagtgctgctctagcatctctcgaccgaggaac 600
Qy 601 ctggagactgactcatctgtcgccctctctgtctagagcgggcgcccaagaatgttgagc 660
Db 601 ctggagactgactcatctgtcgccctctctgtctagagcgggcgcccaagaatgttgagc 660
Qy 661 cggagcaatgaacccctctcgagcgacagcagcccccgccgtgtctacccctctcccg 720
Db 661 cggagcaatgaacccctctcgagcgacagcagcccccgccgtgtctacccctctcccg 720
Qy 721 gcaactgacgagatgaaagcgttgaaagcgttagccgcccagagaccaccccgagacac 780
Db 721 gcaactgacgagatgaaagcgttgaaagcgttagccgcccagagaccaccccgagacac 780
Qy 781 ttggagcaagagatctgtgtcgaagtgctgtgctcgaagttcgaagttgaatgtgagcc 840
Db 781 ttggagcaagagatctgtgtcgaagtgctgtgctcgaagttcgaagttgaatgtgagcc 840
Qy 841 atctctggagatcttgctctgtatgatagtcgaggaagaaagaaatctctcgagaaactc 900
Db 841 atctctggagatcttgctctgtatgatagtcgaggaagaaagaaatctctcgagaaactc 900
Qy 901 tacttgagactgagactccagactccatgaaggggctgtctcgaggtctcatgtgcaaccaact 960
Db 901 tacttgagactgagactccagactccatgaaggggctgtctcgaggtctcatgtgcaaccaact 960
Qy 961 ggcattccaccccttgccgctctgtcacatctctctgtactaacccctcaccttgacatc 1020
Db 961 ggcattccaccccttgccgctctgtcacatctctctgtactaacccctcaccttgacatc 1020
Qy 1021 ttctctgtatcaaatgttgaaagagtgtctcagcaagggagacataagtgtgtgtgag 1080
Db 1021 ttctctgtatcaaatgttgaaagagtgtctcagcaagggagacataagtgtgtgtgag 1080
Qy 1081 ccctacatgtgttgaaagagtgtgacacagcacaagaagaaagaaagcttgaaagagtgtg 1140
Db 1081 ccctacatgtgttgaaagagtgtgacacagcacaagaagaaagaaagcttgaaagagtgtg 1140
Qy 1141 cggctggcgagcgagcaggtctctgcaccccgctggcgctacagcagatccctctgcctg 1200
Db 1141 cggctggcgagcgagcaggtctctgcaccccgctggcgctacagcagatccctctgcctg 1200
Qy 1201 aagggcggtcactcttgcccaacactgtgagcagcgtctggagcagcttgagccggagactgac 1260
Db 1201 aagggcggtcactcttgcccaacactgtgagcagcgtctggagcagcttgagccggagactgac 1260
Qy 1261 tcgagagcgagcgagcgagcagcctgagacagacagcagcgcctctcgaggggccccagagccgg 1320
Db 1261 tcgagagcgagcgagcgagcagcctgagacagacagcagcgcctctcgaggggccccagagccgg 1320
Qy 1321 gcgagtaagtggagcagcagcctgcagctctctgtgctctcgctccagcagcagctactgtc 1380
Db 1321 gcgagtaagtggagcagcagcctgcagctctctgtgctctcgctccagcagcagctactgtc 1380
Qy 1381 aaaaactctttaagcagaagagcttgagcagactcagttgacagagacactctcaagttccctg 1440
Db 1381 aaaaactctttaagcagaagagcttgagcagactcagttgacagagacactctcaagttccctg 1440
Qy 1441 gctgacatgagagcgagcgtctccctgtgctgagcagcagctccctgtgagctgagccagctc 1500
Db 1441 gctgacatgagagcgagcgtctccctgtgctgagcagcagctccctgtgagctgagccagctc 1500
Qy 1501 aagatcgacattctctcggtctctgtaaaatcccaactctgctctccctgagctgtctt 1560
Db 1501 aagatcgacattctctcggtctctgtaaaatcccaactctgctctccctgagctgtctt 1560

Db 1501 aagatcgacattctctcggtctctgtaaaatcccaactctgctctccctgagctgtctt 1560
Qy 1561 calataagccctacccgagaccacagggcgccgcccacaaagagatcttgatctccc 1620
Db 1561 calataagccctacccgagaccacagggcgccgcccacaaagagatcttgatctccc 1620
Qy 1621 gcccggaaggtctatgcccccatcacagctacaggaacctgtgtacgtgtacccgac 1680
Db 1621 gcccggaaggtctatgcccccatcacagctacaggaacctgtgtacgtgtacccgac 1680
Qy 1681 agcctcaacttcaagcagccgagggctcgtgtgcgaacactgtgtgtgaggtgcaatgac 1740
Db 1681 agcctcaacttcaagcagccgagggctcgtgtgcgaacactgtgtgtgaggtgcaatgac 1740
Qy 1741 atgaaagcgagagaccaccaagcgtctgcggtlcatctttgtgcaagtccagctgagat 1800
Db 1741 atgaaagcgagagaccaccaagcgtctgcggtlcatctttgtgcaagtccagctgagat 1800
Qy 1801 gaattaccgcgagagccttcaacacgggtgtgtctacaaatgctcccgagttctac 1860
Db 1801 gaattaccgcgagagccttcaacacgggtgtgtctacaaatgctcccgagttctac 1860
Qy 1861 gagaggttcaagctgtgcatctctcagcctgtgtgacagagaacacatcaactgtgttcaac 1920
Db 1861 gagaggttcaagctgtgcatctctcagcctgtgtgacagagaacacatcaactgtgttcaac 1920
Qy 1921 ttctaccaaagtcaagctgtgcaagcccgccgagcagctcctctgagagaaacccgtggactt 1980
Db 1921 ttctaccaaagtcaagctgtgcaagcccgccgagcagctcctctgagagaaacccgtggactt 1980
Qy 1981 acttgatccacactgtctgacagcaagcgcgctctgagagacggccccctctgtctccagtg 2040
Db 1981 acttgatccacactgtctgacagcaagcgcgctctgagagacggccccctctgtctccagtg 2040
Qy 2041 tctgtgagccagccgcccagctatctcggtctcaacccgagtgtgagctctccgggc 2100
Db 2041 tctgtgagccagccgcccagctatctcggtctcaacccgagtgtgagctctccgggc 2100
Qy 2101 atgctgaggtgtgagcaggttcaaaaggggtgttgaggttgaaagcagcggtgtctct 2160
Db 2101 atgctgaggtgtgagcaggttcaaaaggggtgttgaggttgaaagcagcggtgtctct 2160
Qy 2161 gtgcacccccagagacccctcactctgagacaatctctcaacccctgtgtgacgtctctgagag 2220
Db 2161 gtgcacccccagagacccctcactctgagacaatctctcaacccctgtgtgacgtctctgagag 2220
Qy 2221 ggaagccttcccatctccggtcgaagagacatgtgtctgagcagggcaagctgtgagcagag 2280
Db 2221 ggaagccttcccatctccggtcgaagagacatgtgtctgagcagggcaagctgtgagcagag 2280
Qy 2281 ctgagggcagactctgagcagctgtgcctgtgcagcccgcaaacctctgttgccctctc 2340
Db 2281 ctgagggcagactctgagcagctgtgcctgtgcagcccgcaaacctctgttgccctctc 2340
Qy 2341 caccacgtgtctgagacaagctcgtgtgctgtgtlcatcaaggcccccgatcatcagttgagcag 2400
Db 2341 caccacgtgtctgagacaagctcgtgtgctgtgtlcatcaaggcccccgatcatcagttgagcag 2400
Qy 2401 attgtgaacctgtggcggtgagagccttgaaagaaatgagcccaatgtgtgagcctgttcaac 2460
Db 2401 attgtgaacctgtggcggtgagagccttgaaagaaatgagcccaatgtgtgagcctgttcaac 2460
Qy 2461 cggagccttgagagcagcccaagatgcccgcggtlcaactgtgcccagaactgtgctcctacgtc 2520
Db 2461 cggagccttgagagcagcccaagatgcccgcggtlcaactgtgcccagaactgtgctcctacgtc 2520
Qy 2521 caactagccttctgacctctgtgacactgagcccaagcctcccgatgtggggccccctccagtg 2580
Db 2521 caactagccttctgacctctgtgacactgagcccaagcctcccgatgtggggccccctccagtg 2580
Qy 2581 acagtgcaagcgtgcaacatctgcccgtgtgctctgtgtgcgccccgaaagcctctaactgagc 2640
Db 2581 acagtgcaagcgtgcaacatctgcccgtgtgctctgtgtgtgcgccccgaaagcctctaactgagc 2640
```

Db 2581 ACAGTCAGGCTTCACACTGAGCCGTGGCTGTGGTGGCCGCCCAAGCCTTACTGTGGC 2640  
QY 2641 cgttccaagaatcaagcaagcaaaccttgaacctgcctgcctgagcccttgctctgtgat 2700  
Db 2641 CGTTCCAAAGAGCTAGCAGCAGCAACCTTGACTGTGGCTGGCCCTGGCTGTGGAT 2700  
QY 2701 gacgaagttcccgatcctctgacgaagctgcttcaagaagagcttgctctgtgatg 2760  
Db 2701 GAGGAGGTTCCCGCATCTGGCCACAGAGCTTCACAGGAGAGCTGGCTTGGAGTGG 2760  
QY 2761 gttgtcaagcagatgtgcgttaacgagagcaacctcaagcaagctggtctcttccag 2820  
Db 2761 GTGTGAGAGCAGCTGCTTACCGCAGGCACTCTCCACAGCAGCTGGTCTTCTTCCAG 2820  
QY 2821 ctcaatgtgaagatgtaaggcgtgcaactgctgcttgacgaagactagacaacccgac 2880  
Db 2821 CTGATGTTGAAGATGTAGCGGTGACACTGCTGTGGCCAGCAGCTAGACACCCGCC 2880  
QY 2881 aagctgcgttcctccggaagcttctctggaagacataactgcttggctgtgacctgtggc 2940  
Db 2881 AACCTCGCTTCCCGGAGGCTTCTGAGCAGACATCAGCTTGGTGGCTGTGGGAC 2940  
QY 2941 ctggaagtcataaccgctgttcaacaagaatgtggaagctgacgaagcaacctcaagccagc 3000  
Db 2941 CTGGAGGTATCACCCGCTGTCCACAAGGATGTGAGACTGGCCGAGACCTCAACGCCAGC 3000  
QY 3001 ctggtcttctctcaatgtaacctgtgctctgtgtagacggggcttgtcttcaagctg 3060  
Db 3001 CTGGCTTCTTCTCAGTACACTTCTGTCCCTGGTGGAGCGGGGCTTGTCTTCAAGCTG 3060  
QY 3061 gtccggagccacataagaaggttgacagcggtctccagtgctccctaatcaacgaagc 3120  
Db 3061 GTCCGGGCCCCATPACAGAGGAGGAGGCGGAGCTCCACTGCTCCCTAATCCAGAGGC 3120  
QY 3121 ctgtctgaacctgagcaatggaatcacccgcatcctgtgacgaacgaagcaactagtgacc 3180  
Db 3121 CTGCTGACCCCTGAGCATGGAATTACACCGCATTCCTGTGAGCGCAGAGCATACGTGAGAC 3180  
QY 3181 ctcaacctccctgtgctgccccctgttaacctcaagctgacctccccctctgtgtctcc 3240  
Db 3181 CTCACCTCCCTCCGTGCGCCCTGTACCTCCAGCTCCCTCCCTCCCTGTGTGTCCTCC 3240  
QY 3241 accaactccaaaggtccaaactctcagccaaagcccggaagcccaaggttgacgaagatg 3300  
Db 3241 ACCAAGCTCCAGAGCTCCACTTCTCCAGCCAAAGCCCGGAGCCCAAGGTGACAGCAATG 3300  
QY 3301 ttcgaactgagtgagcaatcccgagcagcaacttctctagcttgagctcctctgctgaagag 3360  
Db 3301 TTGGAAGTGAAGTGAAGCAATTCGCGAGCAGCACTTCTAGCTGGGCTCCTGCTGAGCGAG 3360  
QY 3361 ctggaactgagctctgaactgagaggttgaaagggaatctcctgttgacaagaagaagccatc 3420  
Db 3361 CTGGGCACTGGCCCTTGACACCTGAGGCTGAAGGGGCAATTCCTGTGTGCAAGAAAGGCATTC 3420  
QY 3421 agtgcgtgtcaagcgtctgatagtgacatagaactgaacccccgcgtacagccgaagccatc 3480  
Db 3421 AGTGCGTGTGACAGCTGTGATGTGGCAATGACACTGACCCCGCTAGCCGCAAGGCACT 3480  
QY 3481 gttgaaggtcgtgtgacgaagctgtactcgtgcaactgcttgcgattgacagggataactgt 3540  
Db 3481 GTTGAAGGCTGTGTGGCCAGAGCTGTACTGTCACACTGTTTCATTGACAGGATACCTTG 3540  
QY 3541 ccaagctgcatgtaacttctgtagggcccaaggtcaagcgttcaagactggtcccaatgctt 3600  
Db 3541 CCAGGCTGTGATTAATTTGCTGAGGGCCCAAGGTCAAGGCTCAAGATGTGCTCAATAGCTT 3600  
QY 3601 gactcagacacayaagcgaaaggagacatgtgcgttaccatacaaacccctctgtgacatg 3660  
Db 3601 GACTCAGACACAAAGGCGAAGGGGACATTTGGGGTACATCAACCCCTCTGTGGGCAATG 3660  
QY 3661 gccattgctgtgtgccccctagccccctggtcccgagcagaatctccgaaggccaca 3720  
Db 3661 GCCATTGCTGTGGTGGCCCTTAGGCCCTGGGCTCCGCGGCGCAGCATCTCCAGGGGCCACCA 3720

QY 3721 aaggtctctgcgcaagcgtgtgccccctctctgtgtaagtaagccggaacctgtgtggtgt 3780  
Db 3721 AAGGCTTCTCCGCAAGGCTGTGGCCCTCTGCTGTGATCAAGCCGAGACTGTGGGCTGTG 3780  
QY 3781 gttgctgtgtgtgtgtgaaaaacacccgaagccgggctccttgaaagcgttgagccactgacctg 3840  
Db 3781 GTGCTGTGTGTGTGTGAAAACACCGAGCCGGGCTCCTGTGAGCGCCTGTGGCACTGAGACTG 3840  
QY 3841 aacctccccccagctgggaagctgtgtgactgtgtgaacttgccttgctgctgttgaag 3900  
Db 3841 AACCTCCCCCAGCTGGAGCGTGTGTGACTGTGTAACCTTTCCTGGGCTGTGGAG 3900  
QY 3901 tacaaggggaaaaagcccttgaacgatacaaacgctcacaatcaaaaaatctctggat 3960  
Db 3901 TACAAGGGGAAAAAGCCCTTGAACGATCAACAGCCTCAATCAAAAAATCTCTGGAT 3960  
QY 3961 atgaaggcgsgctaaagggaagccattctgtgtacatctggaagcttgacaagaatggtt 4020  
Db 3961 ATGAAGCGCGGCTTAGAGAAAGCCATTCTGGGTACATCGAGGCTCGACAAAGAAATGTT 4020  
QY 4021 cggcgaaatcgtgagagagagccgcttgggaatccagaagaactgtgcgttgcggaagagc 4080  
Db 4021 CGGGAAGTGTGAGAGGAGCGCTTTGGGAATCCAGAGAAAGTGGCTGGCGGAAGAGC 4080  
QY 4081 gtcaacaactgtgaagcaaacctcaagccgctgtggaacagaacaaagatgaatggaacac 4140  
Db 4081 GTCAACACTGTGAAGCAAACTCAGACCGCGGTGACAAAGAACCAAGATGAATGAAGAAC 4140  
QY 4141 gaagcctgtgtgaaaggaaactgtgcaacggaggcaagcctagtgtgtcttggaacacatg 4200  
Db 4141 GAGGCTTGGTGTGAAGGGAAGCACTGTGGCAACCGAGCAAGCCTAGTGGTTCTGAGACACTG 4200  
QY 4201 gagatcatcgtgcaagcgtgtatgtcttccagaagcccggaagagcgtcttgtygggagatg 4260  
Db 4201 GAGATCATCTGTGCAAGCGGTGATGCTTTCAAGACCCGGGAGACGTTGGGGGCGAGTG 4260  
QY 4261 ctgaaggtgtgtgtgtacagccttggtgcaagtgccccagatgtcctctcttgcaagcatgac 4320  
Db 4261 CTGAAGGTGTGTGTATAGCTGTGGGAGTGCCCAAGATGCCCCCTTCTTGAAGCATAGGC 4320  
QY 4321 ctggaaccccaagaggccctgtgttccaaagtctccggaagctgctgttccgaaggagacag 4380  
Db 4321 CTGGGCAACCAAGAGGCGCTTGTGTCCAAAGTCCCGGAGCTGCTGTTCGAGAGAGACAG 4380  
QY 4381 gagctgtgtgcgaactgtgtcctgaagctcctacgaacatgtgtgacgcgcatcgaacc 4440  
Db 4381 GAGCTGTGTGCGGCACTGTGTGCTGAGGCTCCTACGACACTGTGGCAGCCCATAGCACC 4440  
QY 4441 atccgcaagcagcgaagcctgtgcttaacctgtctatggaacgaagaacttcgaagatgcgc 4500  
Db 4441 ATCCGACGCAAGCCAGCGGCTGTGCTGTACTGTCTATGCAAGCAAACTTGAGATGGC 4500  
QY 4501 cacaacttgcgcgtgtgaaagtacaggtcaacatgtctctctcctgcctggttggtgagc 4560  
Db 4501 CACAAGCTTGGCCCGGTGAAGATGACAGTGCATGTCTGTGCTGTCGATGGGGAGC 4560  
QY 4561 acgcaagaactcagtgaaagagcacttgagcgttcaactcaaaacatcctcaactatgct 4620  
Db 4561 ACGCAGAAGCTTCAAGTGAAGAGCACCTGTGAGCTTCACTCAAAAACATCTCACCTATAGCT 4620  
QY 4621 gagaggaacatgagggctgtgcggaagacaccttgcgaagcaggtgtccagggcctgatatgtc 4680  
Db 4621 GAGGAGGACATGGGGCTGTGGGAGCAAGCTTGTGCAAGGAGGCTGACGAGACTGATATTTC 4680  
QY 4681 aaactgtcacatgatacttgaacgagacaggttgaaaga tgaaggaaacacgaaggaggaacctgag 4740  
Db 4681 AACGTGCAGATGATCTGACGAGACAGGTGAAGATGAAGGAACCAAGAGAGACCTGTAG 4740  
QY 4741 atgtctcatcgacctcaatgataagaattgcccggggtctaaccaagggtctcaacggacctctcg 4800  
Db 4741 ATGCTCATCGACCTCATATGTACGAATTTGCCGGGGCTTACCAAGGGCTTACCGGAGACTTGG 4800

QY 4801 ctgacctgtgtgcagacaatgcccggagagacgcgcgcgcgtgtgccaaccacgcgcgcgcgc 4860  
|||||  
Db 4801 ctgacctgtgtgcagacaatgcccggagagacgcgcgcgcgtgtgccaaccacgcgcgcgc 4860  
QY 4861 gccacgtgcatgtgtgcacgcgcgcgc 4886  
|||||  
Db 4861 gccacgtgcatgtgtgcacgcgcgcgc 4886

RESULT 3  
BC008335 2825 bp mRNA linear PRI 12-JUL-2001  
LOCUS Homo sapiens, clone IMAGE:3506202, mRNA, partial cds.  
DEFINITION BC008335  
ACCESSION BC008335.1 GI:14249910  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2825)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (25-MAY-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC) Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: DCTD/DRP  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@gscc.bc.ca](mailto:info@gscc.bc.ca)  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
Susana Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
Sen Lee, Victor Ling, Carrie Mathewson, Candace McLeay, Steven  
Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline  
Schell, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,  
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,  
George Yang, Scott Zuyderduyn, Marco Marra.

REMARK  
COMMENT  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAL Plate: 23 Row: d Column: 12.  
Location/Qualifiers  
1. 2825  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3506202"  
/tissue\_type="kidney, renal cell adenocarcinoma"  
/clone\_lib="NIH\_MGC\_14"  
/lab\_host="DH10B-R"  
/note="Vector: pOT7"  
<1. 2585  
/codon\_start=3  
/product="Unknown (protein for IMAGE:3506202)"  
/protein\_id="AAH08335.1"  
/db\_xref="GI:14249911"  
/translation="ORSLASMLSDTDEGEDIACTINPSVAMAIAGAPLAPGRASRASI  
SOGPTPASAGCALSAESSRTLLACVLWKLNTBEPALLORMATIDTLTOLGLDILLY  
LCIAPEYKGRKAFERINSLTFFKSLDMKARLEAIIIGTIGAROMVRSRSPFG  
PENVRMRKSVTHMKOTSDRVDKTDMEHEALVEGNIAFEASLVYLDLLETITCYML  
SEARRESVGAIVLKVLYSISGNOSALFLOHGATATRALVSKPELLFEEDBELCADC  
LRLRRSGSRISTIRTHASASLYLMMKONFETGHNFAFKMKQVTVSLSLVGTQNF  
EBHRLRSKTLITVAEDMGLDSTFAEQVODLNMNLHIIITDVTKMEHEDPMLI  
DLMYRLARGYSGSPDLRLTWMQNAKNAEONHAEQNAHAALVAEYLALEDEH  
RHLPVQCVSFQNISSNVLEESAISDDISLSPDESGCKHFEELGLVGLDAAYFT  
MGLVEAVNEVYKNLIPITLAEHRDYKKLAAYHGIKQLEAFTRIMHOSGMEVYFYFR  
VGYGAHFGDDEGEFYVKPEISITKLAEISHLLEFITEREDDVEYLTKNSNPVKS

FEATURES  
source

CDS

BASE COUNT 641 a 835 c 827 g 522 t  
ORIGIN  
Query Match 43.9%; Score 2797.8; DB 9; Length 2825;  
Best local Similarity 99.9%; Pred. No. 0; Mismatches 2; Indels 0; Gaps 0;  
Matches 2799; Conservative 0;  
QY 3572 gtcaagcgtcaaacgactgtgcctcaatgtcttgaactcagacacagaagcgaagggagacatg 3631  
|||||  
Db 1 gtcagcgggtcaagactgtgcctcaatgtcttgaactcagacacagaagcgaagggagacatg 60  
QY 3632 cgggtaccatcaacccctctgtgccaatgtgccaatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3691  
|||||  
Db 61 cgggtaccatcaacccctctgtgccaatgtgccaatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 120  
QY 3692 cccgggccaagatctcccaaggggccaagcggcttctgcgcaagcgtgtgtgtgtgtgtgtgtgt 3751  
|||||  
Db 121 cccgggccaagatctcccaaggggccaagcggcttctgcgcaagcgtgtgtgtgtgtgtgtgtgt 180  
QY 3752 ctgaagtcgaagc 3811  
|||||  
Db 181 ctgaagtcgaagc 240  
QY 3812 cgcctctgaagc 3871  
|||||  
Db 241 cgcctctgaagc 300  
QY 3872 tgcctgtaccctgt 3931  
|||||  
Db 301 tgcctgtaccctgt 360  
QY 3932 acagcctcaccatcaaaaaaattctgtgatalgaagaagcgcgcgcgcgcgcgcgcgcgcgcgc 3991  
|||||  
Db 361 acagcctcaccatcaaaaaaattctgtgatalgaagaagcgcgcgcgcgcgcgcgcgcgcgcgc 420  
QY 3992 gtaccatgtgaactgtgcaagaagaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4051  
|||||  
Db 421 gtaccatgtgaactgtgcaagaagaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480  
QY 4052 atccagaagaagcgt 4111  
|||||  
Db 481 atccagaagaagcgt 540  
QY 4112 tggacaagaaccaagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 4171  
|||||  
Db 541 tggacaagaaccaagatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 600  
QY 4172 agtgaagcctgt 4231  
|||||  
Db 601 agtgaagcctgt 660  
QY 4232 aagcccggaagagcgtctgt 4291  
|||||  
Db 661 aagcccggaagagcgtctgt 720  
QY 4292 cccaagatgtcctctctctgtgacatgtgctgtgcaaccacagaagggcctgtgtgtcaagt 4351  
|||||  
Db 721 cccaagatgtcctctctctgtgacatgtgctgtgcaaccacagaagggcctgtgtgtcaagt 780  
QY 4352 tccggagatgtctgt 4411  
|||||  
Db 781 tccggagatgtctgt 840  
QY 4412 taagcacatgt 4471  
|||||  
Db 841 taagcacatgt 900



Oy	4472	tgccatctcgacgaacttcgagatccgycaccaactttgccgttgyagatgacaggtca	4531
Db	901	tgctcaatgcacagaaactttcgaaatcgccacaaactttggcccggtgaatctcaggtca	960
Oy	4532	ccatgctctctgctccctggtggtggaagacgcgaacttcagtgaagagacccttcgac	4591
Db	961	ccatgctctctgctccctggtggtggaagacgcgaacttcagtgaagagacccttcgac	1020
Oy	4592	gttcaactcaaaacatcctccactatgctgagggagagacatgggctgcyggacagacct	4651
Db	1021	gtttcactcaaaacatcctccactatgctgagggagagacatgggcttcgggagacagacct	1080
Oy	4652	tcgcaagagcaggttcgaagaccctgcatggttcaaaccttcgacatgctcgaagacgttga	4711
Db	1081	tcgcagacgacggtccagagacctgcatggttcaaaccttcgacatgctcgaagacgttga	1140
Oy	4712	agatgaagagaacaccagaggaacctgagatgctcatcgacctatgtacagaattgccc	4771
Db	1141	agatgaagagaacaccagaggaacctgagatgctcatcgacctatgtacagaattgccc	1200
Oy	4772	ggaggtccacaggtctacacggacccttcggtctgacctgtgttcgaagaatgcccgggaacg	4831
Db	1201	gggggcttaccaggtctacacggacccttcggtctgacctgtgtgttcgaagaatgcccgggaacg	1260
Oy	4832	acgcgagctctggtjoaacaccacgcccggcccgccagtgatgagtgcagcggccgcccctcg	4891
Db	1261	acgcggagctctggtjoaacaccacgcccggcccgccagtgatgagtgcagcggccgcccctcg	1320
Oy	4892	tggtctgagttacctcgccctgtgcttgagagaccagcgcacccttcgggtggtctgcttcc	4951
Db	1321	tggtctgagttacctcgccctgtgcttgagagaccagcgcacccttcgggtggtctgcttcc	1380
Oy	4952	tcacgaacatctcaatccaaagtctgtagagagttccgcacatctccgacgaacatccgtctgc	5011
Db	1381	tcacgaacatctcaatccaaagtctgtagagagttccgcacatctccgacgaacatccgtctgc	1440
Oy	5012	ccgacgagagaggtctctgtctccgggaagcaacttaactgagcttgggctgtgtaggtgtgc	5071
Db	1441	ccgagcagagaggtctctgtctccgggaagcaacttaactgagcttgggctgtgtaggtgtgc	1500
Oy	5072	tggaacagcgacgctgcttcaacatgagtggtgcggtcctacgagtcggttgatgaagttct	5131
Db	1501	tggaacagcgacgctgcttcaacatgagtggtgcggtcctacgagtcggttgatgaagttct	1560
Oy	5132	acaagaacctcaatcccaactctgtgaagcccaacgctgactcaagaagctgtgcgcggtgc	5191
Db	1561	acaagaacctcaatcccaactctgtgaagcccaacgctgactcaagaagctgtgcgcggtgc	1620
Oy	5192	acggaacaactcaggaagaccttcaaccaagatcatgaccagaagttccgctctggagacgcg	5251
Db	1621	acggaacaactcaggaagaccttcaaccaagatcatgaccagaagttccgctctggagacgcg	1680
Oy	5252	tggtcggagcgtatctccgctgtggtcttcaagcgcccaacttcggtgagcttgatgagc	5311
Db	1681	tggtcggagcgtatctccgctgtggtcttcaagcgcccaacttcggtgagcttgatgagc	1740
Oy	5312	agggaattgtgtlaaagaagacatcgatccgaagcttggcaagatctcaaccggctgg	5371
Db	1741	agggaattgtgtgtlaaagaagacatcgatccgaagcttggcaagatctcaaccggctgg	1800
Oy	5372	agggaattctacccgagaggaatttggtgcagacgtctgtagaatttcaaaagacttllac	5431
Db	1801	agggaattctacccgagaggaatttggtgcagacgtctgtagaatttcaaaagacttllac	1860
Oy	5432	ctgtggaacaagtccaagcttgactacacaaagctcaatccagatcatcgtatgtggaac	5491
Db	1861	ctgtggaacaagtccaagcttgactacacaaagctcaatccagatcatcgtatgtggaac	1920
Oy	5492	ctgactctgtatctcagaagcttcaagaaccgggttgactactttgacccgcaactctggc	5551
Db	1921	ctgactctgtatctcagaagcttcaagaaccgggttgactactttgacccgcaactctggc	1980
Oy	5552	tttgcacatctctgttctgcacgcggttcaacgcggaattggcgcgacacagggagactgc	5611

D	b	1981	TTCGACATTCTTGTTCGCAGCCCGTTCAAGCCGGATGGGGCAGCACGGGGAGCTGC		2040
Oy		5612	cgcagcaacacaaagcgttaagaacctgcttcgaaccgcccgctcccccataataga		5671
D	b	2041	CCGGACAACACAAACCCTGAAGAACCTGGTCAGCACCGAACAACGCCCTTCCCCTAACATCAAGA		2100
Oy		5672	ctgcattccgtgtgtgccaccaacggaggagaacgctgcgtcacgcagtcagtgtagtggccaicg		5731
D	b	2101	CTCCGATCCGCTGTGTGCCACCGGGAGGAGAACGGTGCTGAGGCCAGTGTGAGAGTGGCCATCG		2160
Oy		5732	aggacaatgcagaagaagacacacggaagtgtgctctttgccaacgagcagaccaccagatg		5791
D	b	2161	AGGACATCCACAAGAGAGACACGGAGCTGGGCTTTGGCCACCGACGACGACCCACCAATG		2220
Oy		5792	ctaagatctcaaatgatgtgtttaaggaactgtfaaggccccacacgltaaccaagtlcccc		5851
D	b	2221	CTTAACATCTTACAGATGCTGCTTCAAGGCTCTGTAGGGCCACACGTTAACACAGGGTCCCC		2280
Oy		5852	tggaggtgcccaggtgtlittlaogaagalcccggaaagaccacgaactcttcggcatc		5911
D	b	2281	TGGAGGTGGCCACGAGTGTTTTAGAGAGATCCCGAAGAACCCCAACTCTTCGGGATC		2340
Oy		5912	acaaaattgcggtgcgtcgtctcaaggaactcttcaaanaattgagagatgcgttcgga		5971
D	b	2341	ACAACAATAATGGGGCTTCGCTTCAAGACCTTCTGCAGAAAATGTGAGGATTCGCTGGCGGA		2400
Oy		5972	aaaaaagaccctgaatlbggcccgaaccagaagagtatccaccgtlgagcttgagacgcaact		6031
D	b	2401	AAAAATAAGGCCCTGATGTGGCCGGACCCAAGAGAGATCACCGTAGCTGTGAGCGCAACT		2460
Oy		6032	actgcgcgctgcggaagagctctgcagcccctgcgttacccaaagcgcgttggcccaagttatg		6091
D	b	2461	ACTGCGCGCTGGGAGAGGCTTCGACGCCCTTGCTTAACCAAGCGCTGCCCAAGCTGAAGG		2520
Oy		6092	caaccaccccaaccgagctcagaagaactctcttgaacagagcaagtltcgnaaaggcaagcc		6151
D	b	2521	CACCCACCCACCGCGGCTCAGAGAACTCCTTGAACAGAGCAAATTCGAAAGGAGAGCC		2580
Oy		6152	tcttgagcccaagaagcaacaaagcttgtactatgagaacaacgaacccggggcttaagctgtc		6211
D	b	2581	TCTGACCCCAACAAGACCAAMAAGCTGTACTTGAAGAGACCAAGACCCGGGCTCAGACTTC		2640
Oy		6212	tgtgctgcagaaggagatctgcacctgtgtgcoccaacttgagctgttgggttgaaaccaactgtact		6271
D	b	2641	TGTGCTGCAGAGGAGTCTGCGCCTGGTGGCCACTGTGGCTGTGGSGTGACACACTGTACT		2700
Oy		6272	tggggctgggcccctctgcctcgtgtcccaactctgtfgcgaetgshtgcttcctcccttt		6331
D	b	2701	TGGGGCTGGGCGCTCTGCCCTCTGTGCCCATCTGTGTCACTGATGCTCTTCCCTTTT		2760
Oy		6332	ttaatitlaaatgtgttltatalaagcaaaaaaaaaaaaaaaaaa 6372		
D	b	2761	TTAATTAAAATGTGTTTTATTAAGCAAAAAAAAAAAAAAAAAA 2801		
RESULT 4					
		AXI73233	2149 bp	DNA	linear PAT 03-JUL-2001
LOCUS AXI73233					
DEFINITION Sequence 59 from Patent WO0142295.					
ACCESSION AXI73233					
VERSION AXI73233.1 GI:14598038					
KEYWORDS .					
SOURCE	human.				
ORGANISM	Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
REFERENCE	1 (bases 1 to 2149)				
AUTHORS	Lu, P., Garman,J.D. and Candia,A.F.				
TITLE	Clasp-7 transmembrane protein				
JOURNAL	Patent: WI 0142295-A 59 14-JUN-2001;				
Arbor Vita Corporation (US)					
Location/Qualifiers					

source 1. .2149  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="preliminary human CLASP-7 cDNA sequence"  
CDS <2..1933  
/note="human CLASP-7"  
/protein\_id="CAC43732.1"  
/db\_xref="GI:14598039"  
/translation="LSAREASYLDLKVLYLSGSASALFLHGLATQRLVSKPE  
ELLFEDPEEDLADLCRLRHGSRISITRTHASLTLKQNFIEIHNARVKKMOY  
TMSLSLVTTQNFSEHLRNSLKITLYAEEDMLRSTAEQVQDLMFNILHILTD  
TVKNEHODEPEMLIDLMVRIARGYQSGPDLRLTLQNMAGNAHLEGNHADAOCMVH  
AALVAEYLALLEDRHLPLVGCVSFQNTISNVLESAISDILSPDEGFCGKHETE  
LGVLLEQAAAGYFTMGGLYEAVNEVYKNIPLIEAHNDYKRLIAVHGKLOEAPKTM  
HOSGMEVFEGYFRVGFYGAHFDLDOEFYKRPSTYTKIALEISHRLREBYTERFGD  
DYVELIKDSYPIVDSKILDSOKAYIOITTYVEPFDYELKDRVTFDRNYGLRTLEPCT  
PRTPPGRAGNELPEQHKRKTLLSTIDHAFYIKTRIRVGHREYVLTPEVAIEDMOK  
TRLEAFTEQDDPDAKMLQMLVQGSVPTVNOGLEVAQVLELAETPEDPKLFRHNNKL  
RLCFRDFCKCEDALRNKALIGPDKSEYHRELEERNYCRLEALQPLTLQRLPOLMAP  
TPPGLRNSLNRRASFRKADL"

BASE COUNT 468 a 650 c 625 g 406 t  
ORIGIN

Query Match 33.6%; Score 2140; DB 6; Length 2149;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2143; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4225 ctttcagaagcccgagagagcgcttcgtgggggagcagtcgtgaaggttgctgtcaagcctg 4284  
DB 2 CTTTCAGAAAGCCCGGAGAGCGCTTTGGGGATTGCTGAAGGTGTCTGTACAGCCTG 61  
QY 4285 ggcagtcgccagagatgcccctctctcttcagcattgacctggccaccagagggcccttg 4344  
DB 62 GCGAGTGCACAGAGTCCCTTCTTTCAGCATGGCCTGGCCACCCAGAGGGCCCTTTGTG 121  
QY 4345 tccaaagtcgccgagctgtgtctgaagagagacagagagctgtgtgcgaacctgtgctg 4404  
DB 122 TCCAAAGTTCGCCGAGCTGTCTTCAGAGAGGACACGAGAGCTGTGCCGACCTGTGCTG 181  
QY 4405 aggcctctcagacactgtgacagcgacatcagcaaccatccgacagcagcgccctcg 4464  
DB 182 AGGCTCTCAGACACTGTGGGAGCGGCATACAGACATCCGACGACGCGCCCTCG 241  
QY 4465 ctgtaccctgctatctgacagagaactctgaagatcgccaaacttgcgcgtgtgaagatg 4524  
DB 242 CTGTACTGCTCATCGACAGAACTTCAGATCGGCCCAACTTGGCCGCTGTGAAGATG 301  
QY 4525 caggtcaccatgtctctctgctccctgtgtgggagacagcagaacttcagtgaaagacac 4584  
DB 302 CAGGTCAACATGTCTCTCGTCCCTGTGGGAGACGACGAACTTCAGTGAAGAGCAGC 361  
QY 4585 ctggcagcttcaactcaaaaacatccatcactatgctgaagagacatgaggctgcggagac 4644  
DB 362 CTGCACAGTTCACCAAAACATCCTACCTATGCTGAGGAGGACATGGGCTCGGGAGC 421  
QY 4645 agcaccttcgacagcagcagctcagagacctgttcaacctgcgacacatgacctgaacggac 4704  
DB 422 AGCACTTCGACAGACAGAGCTCAGACTGTATTTCAACCTGCAACATGATCTTGACGAC 481  
QY 4705 acggtgaagatgaagaaacacacagagagacccctgagatgtctcatcgaacctcatgaa 4764  
DB 482 ACGGTGAAGATGAAGAAACACACAGAGAGACCCCTGAGATGCTCATCGACATCATGTACGA 541  
QY 4765 attcccgagggtaccacagggtaccagaccttcggtgacacctgtgttcgaagaatggcc 4824  
DB 542 ATTGCCGGGGCTACCAAGGCTTCACCGGACCTTGCGTGAACCTGTTGCAAGACATGGCC 601  
QY 4825 ggaagacacgagagctggtgcaaacacgcccagggccgcccagtgacatggtgacgagggcc 4884  
DB 602 GGAAGACACGGGAGAGTGGGCAACACACGCGAGGCGCCCAAGTGCATGTGTGACGCGGCC 661

QY 4885 gccctcgtgtgctgagtaacctcgccctgctcgaaggaaccagcgcaacctgcccgtgggtcgc 4944  
DB 662 GCCCTCGTGGCTGAGTACTCGCCCTCGTCAAGACACCCGCAACCTGCGCGTGGCTGCG 721  
QY 4945 gttcccttcagaacatcatcacaagctgactaagagaggttcgcgacatctccgaagaatc 5004  
DB 722 GTTTCCTTCCAGAACATCTTCATCCATATGTGTAAGGAGTCCGCAATCTCCGAGACATC 781  
QY 5005 ctgtccgcgacagagagagctctctgctccggaagacacatcactgaactgagtgagtgta 5064  
DB 782 CTGTGCCCCGACAGAGAGAGCGCTTCGTCGCGGAAGACCTTCATGAGCTGTGGCTGCTA 841  
QY 5065 ggtttgcggagaaacggcagcgcttactcaacatgggcygggtctcagagcggtgaat 5124  
DB 842 GGGTTGCTGGAAACAGGCGCGCGGTACTCAACATGGCGGGCTCTACGAGCGGCTGAAT 901  
QY 5125 gaggctcaagaagacctcatcccatcctgtgaagcccaacgctgaactcaagaagcttgacc 5184  
DB 902 GAGGTCTACAAAGAACCTTCATCCCATCTCGAAGCCCAACCTGACTACAAAGAACTGGCC 961  
QY 5185 gcggtgcacggcnaaactgcaagagagccttcaaccaagatcatalgcaaccagagttccggtcg 5244  
DB 962 GCGGTGCACGCGCAAACTGCAGAGAGCGCTTCACCAAGATCATGACACAGAGTTCGGCTGG 1021  
QY 5245 gagcgctgttcggagagctatctccgctggtggtctcagagggcccaactcgtgtgacctg 5304  
DB 1022 GAGCGCGTGTTCGGGACGCTATTTCCGCTGAGGCTTCTACGGCCCACTTGGGTGACCTG 1081  
QY 5305 gatgaagaagagtttgttacaagaagcatalgcatcagaagcttgcagaagacatcaaac 5364  
DB 1082 GATGAGCAGAGGTTGTGTACAAAGAGCCATCATGATCAGAACTGGGAGAGATCTCACAG 1141  
QY 5365 cggcttgagaagagttctcacaagagagagatttggcgaagagctgttgaagatatacaagac 5424  
DB 1142 CGGCTGAGAGGAGTCTACAGAGAGATTTGGGACGACCTGTTAGATTATCAAAAGAC 1201  
QY 5425 tctaacctctgtggaacaagcttgactcacaagaagccatactcaagatcaagctat 5484  
DB 1202 TCTTACCCTGTGACAGATGCCAAGCTTGACTACAAAGGCTTACATCCAGATCACTGAT 1261  
QY 5485 gtggaacgtaactttgatacctaagcagctcaagagccgggtgacctacttgaacgcaac 5544  
DB 1262 GTGGAAACGTAATTTGATATCTACGAGCTCAAGAGCCGGGTGACCTTGAACCCCAAC 1321  
QY 5545 tatgggtctgcacatccctgttctgaacgcttcaacgcccgaatggggcgcaacggg 5604  
DB 1322 TATGGGTTTGCACATTTCTGTTCTGCACGCCCTTACGCCGAGTGGGCGCCACACGGG 1381  
QY 5605 gaagctgcgagagcaacaagaagcgtlaagacgctgtcgaacagcaacgaccttccctac 5664  
DB 1382 GAGCTGCCGAGCAACCAAGGCTTAAGAGCTGTGACGACCGACACGCTTCCCTTAC 1441  
QY 5665 atcaagactgcacatcctgtgtgtgccacgagggaagagtgctctgaagcgaatggaagt 5724  
DB 1442 ATCAAGACTGCAATCCGTGTGTGCCACCGGAGAGACGCTGTGCGCAATGGAAGGTG 1501  
QY 5725 gccatcgagaacatgcagaagaagacacggaagcttgaccttgcacacgagagagacca 5784  
DB 1502 GCCATCGAGAGACATGCAGAAAGAACACGAGAGCTTGCTTGGCACCGAGAGACCCA 1561  
QY 5785 ccagatgtcgaagatgtataagatgtgtctcaaggctcgtgtgagggcccaacggtgaacag 5844  
DB 1562 CCAAGATGCTAAGATGTCTACAGATGTGTGCTTCAAGGCTGTGTGAGGCCACCGTGAACAG 1621  
QY 5845 ggtcccttgagagtgcccgaagtgttttaagaagagatcccggaagacccaagctcttc 5904  
DB 1622 GGTCCCTTGAGAGTGGCCAGAGTGTTTTATGACAGAGATCCCGAAGACCCCAAGCTTTC 1681  
QY 5905 cggcatcaaaaataattcggtctgtcttcaagagacttctgaagaatgtgagagatgcg 5964  
DB 1682 CGGCATCAACAAATTTGGGCTTGTCTTCAAGGACTTCTCAGAAGAAATGTGAGAGATCG 1741  
QY 5965 ctgcgaaaataagagccctgtgttggtgcggaacgaagaggtataccacgtagactgag 6024



Db	1742	CTGCGAANAATTAAGGCCCTGATTGGCCGAGACCAAGAGGATACACCTGAGCTGGAG	1801	
Qy	6025	cgcaactactgcccgcctgagggaggtcctgcaagccctgtgttaacccagcgcttgcccaag	6084	
Db	1802	CGCAACTACTGCGCTGCGGAGGAGCTCTGAGAGCCCTGCTTACCCAGCGCTGGCCCCAG	1861	
Qy	6085	ctgatgagccaccccccaccccgctcaaggaactcccttgaaacgagcaaatlccggaag	6144	
Db	1862	CTGATGGACACCCACCCACCGGCTCTAGGAACCTCTTGAACAAGAGATTTCCGAAG	1921	
Qy	6145	gcaaacctctgagcccaagaagcaaacagctgtatccagagaaaccagaccgagcctc	6204	
Db	1922	GCAAGACCTGTAGGCCCAACAGAGCAACGATGACTAGAGGAACACAGCACCGGGGCTC	1981	
Qy	6205	agctgtctgtgcgcgagggagtcgtccctgtgtgcccaatgagcgtgtgggtgtaaccaca	6264	
Db	1982	AGCTGTCTGTCTCTGCGAGGGGAGTCTGCCCTGGTGGCCACACGGGCTCTGGGGGTACCCACA	2041	
Qy	6265	ctgtaactgggggcctgagggccctctgcccctgtgtcccaatcgtgtgcaatgagcttccct	6324	
Db	2042	CTGACTGTTGGGGCTGGGCCCTCTGCCCTCTGTGTCCCACTGTGTGTGACATGATCTTCT	2101	
Qy	6325	ccctttttaaattaaatggtttttaaagcaaaaaaaaaaaaaa 6372		
Db	2102	CCCTTTTAAATTTAAATGCTTTTATATAGCAAAAAAAAAAAAAA 2149		
RESULT	5			
LOCUS	AX172874	6828 bp	DNA	linear
DEFINITION	Sequence 1 from Patent WO0142297.			PAT 03-JUL-2001
ACCESSION	AX172874			
VERSION	AX172874.1			
KEYWORDS	GI:14597911			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE	1 (bases 1 to 6828)			
AUTHORS	Lu, P., Garman, J.D. and Candia, A.F.			
TITLE	Clasp-3: transmembrane protein			
JOURNAL	Patent: WO 0142297-A 1 14-JUN-2001;			
	Arbor Vita Corporation (US)			
FEATURES	Location/Qualifiers			
SOURCE	1..6828			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/note="full-length human CLASP-3 CDNA"			
	22..6824			
	/note="human CLASP-3"			
	/codon_start=1			
	/protein_id="CA043726.1"			
	/db_xref="GI:14597912"			
	/translation="MAERRAFAPKISRVAAEVRKQISGGVSGSPOLKLNINIVNIS			
	HTTVPLEAVDPVDLDEYLTHPLANDSGPLRLTEPPDDIEVYVSPDRCRTLSA			
	VPESEMDPHVDCIRSTEDMAIVIRKIRKLGTGPNPNTLDKREKQGLPROVPESS			
	DEADGNSYQDDDDLRKRSIDTDEPGVAGSIFPDLKSLPALLPLNLDRIPEEEL			
	IDRDNDQKSRNREKELFALHSPDEEPIERLSVPIPKHFGQRLVLCISLKEPI			
	ETEPFASIALYDEVKKKISENEYFDLNSOMGLRPHVPAITITLARSALTSTL			
	YPSDQVFLIVLEKVLQDGGCEGCAEPIRSEMDATRNKLEKLSQADQFCQRIAG			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	TSGDACNLTEFRPRTLTVTNPFKQEGDRSLDEDLKFLADMRKRSYLRRLRITQ			
	LKIDISPAENPHYCLPELLOVKLYPDSRYRPTRELLFPARVYVNTYRNLLIT			
	YPSLTFANRQGSANITVKKVQAGDEPSAKIETKSSCSFSEKAYTAAYVHNHR			
	SPDFHEIKVKLPLATLIDHNLHFTFYVNSCOQKNPLPEPVGTWYIPLMLONRK			
	GFQFVSLERKPOAYSVLSEPEVLPGRKAVDNHKGYNVEVAWSIHQDDPYLDFE			
	FALVALDEHLEPVYIGDMRIEMNNLEIKSISALNSDLEPVERLHLDLKLIL			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	HTTVPLEAVDPVDLDEYLTHPLANDSGPLRLTEPPDDIEVYVSPDRCRTLSA			
	VPESEMDPHVDCIRSTEDMAIVIRKIRKLGTGPNPNTLDKREKQGLPROVPESS			
	DEADGNSYQDDDDLRKRSIDTDEPGVAGSIFPDLKSLPALLPLNLDRIPEEEL			
	IDRDNDQKSRNREKELFALHSPDEEPIERLSVPIPKHFGQRLVLCISLKEPI			
	ETEPFASIALYDEVKKKISENEYFDLNSOMGLRPHVPAITITLARSALTSTL			
	YPSDQVFLIVLEKVLQDGGCEGCAEPIRSEMDATRNKLEKLSQADQFCQRIAG			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	TSGDACNLTEFRPRTLTVTNPFKQEGDRSLDEDLKFLADMRKRSYLRRLRITQ			
	LKIDISPAENPHYCLPELLOVKLYPDSRYRPTRELLFPARVYVNTYRNLLIT			
	YPSLTFANRQGSANITVKKVQAGDEPSAKIETKSSCSFSEKAYTAAYVHNHR			
	SPDFHEIKVKLPLATLIDHNLHFTFYVNSCOQKNPLPEPVGTWYIPLMLONRK			
	GFQFVSLERKPOAYSVLSEPEVLPGRKAVDNHKGYNVEVAWSIHQDDPYLDFE			
	FALVALDEHLEPVYIGDMRIEMNNLEIKSISALNSDLEPVERLHLDLKLIL			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	HTTVPLEAVDPVDLDEYLTHPLANDSGPLRLTEPPDDIEVYVSPDRCRTLSA			
	VPESEMDPHVDCIRSTEDMAIVIRKIRKLGTGPNPNTLDKREKQGLPROVPESS			
	DEADGNSYQDDDDLRKRSIDTDEPGVAGSIFPDLKSLPALLPLNLDRIPEEEL			
	IDRDNDQKSRNREKELFALHSPDEEPIERLSVPIPKHFGQRLVLCISLKEPI			
	ETEPFASIALYDEVKKKISENEYFDLNSOMGLRPHVPAITITLARSALTSTL			
	YPSDQVFLIVLEKVLQDGGCEGCAEPIRSEMDATRNKLEKLSQADQFCQRIAG			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	TSGDACNLTEFRPRTLTVTNPFKQEGDRSLDEDLKFLADMRKRSYLRRLRITQ			
	LKIDISPAENPHYCLPELLOVKLYPDSRYRPTRELLFPARVYVNTYRNLLIT			
	YPSLTFANRQGSANITVKKVQAGDEPSAKIETKSSCSFSEKAYTAAYVHNHR			
	SPDFHEIKVKLPLATLIDHNLHFTFYVNSCOQKNPLPEPVGTWYIPLMLONRK			
	GFQFVSLERKPOAYSVLSEPEVLPGRKAVDNHKGYNVEVAWSIHQDDPYLDFE			
	FALVALDEHLEPVYIGDMRIEMNNLEIKSISALNSDLEPVERLHLDLKLIL			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	HTTVPLEAVDPVDLDEYLTHPLANDSGPLRLTEPPDDIEVYVSPDRCRTLSA			
	VPESEMDPHVDCIRSTEDMAIVIRKIRKLGTGPNPNTLDKREKQGLPROVPESS			
	DEADGNSYQDDDDLRKRSIDTDEPGVAGSIFPDLKSLPALLPLNLDRIPEEEL			
	IDRDNDQKSRNREKELFALHSPDEEPIERLSVPIPKHFGQRLVLCISLKEPI			
	ETEPFASIALYDEVKKKISENEYFDLNSOMGLRPHVPAITITLARSALTSTL			
	YPSDQVFLIVLEKVLQDGGCEGCAEPIRSEMDATRNKLEKLSQADQFCQRIAG			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	TSGDACNLTEFRPRTLTVTNPFKQEGDRSLDEDLKFLADMRKRSYLRRLRITQ			
	LKIDISPAENPHYCLPELLOVKLYPDSRYRPTRELLFPARVYVNTYRNLLIT			
	YPSLTFANRQGSANITVKKVQAGDEPSAKIETKSSCSFSEKAYTAAYVHNHR			
	SPDFHEIKVKLPLATLIDHNLHFTFYVNSCOQKNPLPEPVGTWYIPLMLONRK			
	GFQFVSLERKPOAYSVLSEPEVLPGRKAVDNHKGYNVEVAWSIHQDDPYLDFE			
	FALVALDEHLEPVYIGDMRIEMNNLEIKSISALNSDLEPVERLHLDLKLIL			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	HTTVPLEAVDPVDLDEYLTHPLANDSGPLRLTEPPDDIEVYVSPDRCRTLSA			
	VPESEMDPHVDCIRSTEDMAIVIRKIRKLGTGPNPNTLDKREKQGLPROVPESS			
	DEADGNSYQDDDDLRKRSIDTDEPGVAGSIFPDLKSLPALLPLNLDRIPEEEL			
	IDRDNDQKSRNREKELFALHSPDEEPIERLSVPIPKHFGQRLVLCISLKEPI			
	ETEPFASIALYDEVKKKISENEYFDLNSOMGLRPHVPAITITLARSALTSTL			
	YPSDQVFLIVLEKVLQDGGCEGCAEPIRSEMDATRNKLEKLSQADQFCQRIAG			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	TSGDACNLTEFRPRTLTVTNPFKQEGDRSLDEDLKFLADMRKRSYLRRLRITQ			
	LKIDISPAENPHYCLPELLOVKLYPDSRYRPTRELLFPARVYVNTYRNLLIT			
	YPSLTFANRQGSANITVKKVQAGDEPSAKIETKSSCSFSEKAYTAAYVHNHR			
	SPDFHEIKVKLPLATLIDHNLHFTFYVNSCOQKNPLPEPVGTWYIPLMLONRK			
	GFQFVSLERKPOAYSVLSEPEVLPGRKAVDNHKGYNVEVAWSIHQDDPYLDFE			
	FALVALDEHLEPVYIGDMRIEMNNLEIKSISALNSDLEPVERLHLDLKLIL			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	HTTVPLEAVDPVDLDEYLTHPLANDSGPLRLTEPPDDIEVYVSPDRCRTLSA			
	VPESEMDPHVDCIRSTEDMAIVIRKIRKLGTGPNPNTLDKREKQGLPROVPESS			
	DEADGNSYQDDDDLRKRSIDTDEPGVAGSIFPDLKSLPALLPLNLDRIPEEEL			
	IDRDNDQKSRNREKELFALHSPDEEPIERLSVPIPKHFGQRLVLCISLKEPI			
	ETEPFASIALYDEVKKKISENEYFDLNSOMGLRPHVPAITITLARSALTSTL			
	YPSDQVFLIVLEKVLQDGGCEGCAEPIRSEMDATRNKLEKLSQADQFCQRIAG			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	TSGDACNLTEFRPRTLTVTNPFKQEGDRSLDEDLKFLADMRKRSYLRRLRITQ			
	LKIDISPAENPHYCLPELLOVKLYPDSRYRPTRELLFPARVYVNTYRNLLIT			
	YPSLTFANRQGSANITVKKVQAGDEPSAKIETKSSCSFSEKAYTAAYVHNHR			
	SPDFHEIKVKLPLATLIDHNLHFTFYVNSCOQKNPLPEPVGTWYIPLMLONRK			
	GFQFVSLERKPOAYSVLSEPEVLPGRKAVDNHKGYNVEVAWSIHQDDPYLDFE			
	FALVALDEHLEPVYIGDMRIEMNNLEIKSISALNSDLEPVERLHLDLKLIL			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	HTTVPLEAVDPVDLDEYLTHPLANDSGPLRLTEPPDDIEVYVSPDRCRTLSA			
	VPESEMDPHVDCIRSTEDMAIVIRKIRKLGTGPNPNTLDKREKQGLPROVPESS			
	DEADGNSYQDDDDLRKRSIDTDEPGVAGSIFPDLKSLPALLPLNLDRIPEEEL			
	IDRDNDQKSRNREKELFALHSPDEEPIERLSVPIPKHFGQRLVLCISLKEPI			
	ETEPFASIALYDEVKKKISENEYFDLNSOMGLRPHVPAITITLARSALTSTL			
	YPSDQVFLIVLEKVLQDGGCEGCAEPIRSEMDATRNKLEKLSQADQFCQRIAG			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	TSGDACNLTEFRPRTLTVTNPFKQEGDRSLDEDLKFLADMRKRSYLRRLRITQ			
	LKIDISPAENPHYCLPELLOVKLYPDSRYRPTRELLFPARVYVNTYRNLLIT			
	YPSLTFANRQGSANITVKKVQAGDEPSAKIETKSSCSFSEKAYTAAYVHNHR			
	SPDFHEIKVKLPLATLIDHNLHFTFYVNSCOQKNPLPEPVGTWYIPLMLONRK			
	GFQFVSLERKPOAYSVLSEPEVLPGRKAVDNHKGYNVEVAWSIHQDDPYLDFE			
	FALVALDEHLEPVYIGDMRIEMNNLEIKSISALNSDLEPVERLHLDLKLIL			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	HTTVPLEAVDPVDLDEYLTHPLANDSGPLRLTEPPDDIEVYVSPDRCRTLSA			
	VPESEMDPHVDCIRSTEDMAIVIRKIRKLGTGPNPNTLDKREKQGLPROVPESS			
	DEADGNSYQDDDDLRKRSIDTDEPGVAGSIFPDLKSLPALLPLNLDRIPEEEL			
	IDRDNDQKSRNREKELFALHSPDEEPIERLSVPIPKHFGQRLVLCISLKEPI			
	ETEPFASIALYDEVKKKISENEYFDLNSOMGLRPHVPAITITLARSALTSTL			
	YPSDQVFLIVLEKVLQDGGCEGCAEPIRSEMDATRNKLEKLSQADQFCQRIAG			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	TSGDACNLTEFRPRTLTVTNPFKQEGDRSLDEDLKFLADMRKRSYLRRLRITQ			
	LKIDISPAENPHYCLPELLOVKLYPDSRYRPTRELLFPARVYVNTYRNLLIT			
	YPSLTFANRQGSANITVKKVQAGDEPSAKIETKSSCSFSEKAYTAAYVHNHR			
	SPDFHEIKVKLPLATLIDHNLHFTFYVNSCOQKNPLPEPVGTWYIPLMLONRK			
	GFQFVSLERKPOAYSVLSEPEVLPGRKAVDNHKGYNVEVAWSIHQDDPYLDFE			
	FALVALDEHLEPVYIGDMRIEMNNLEIKSISALNSDLEPVERLHLDLKLIL			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	HTTVPLEAVDPVDLDEYLTHPLANDSGPLRLTEPPDDIEVYVSPDRCRTLSA			
	VPESEMDPHVDCIRSTEDMAIVIRKIRKLGTGPNPNTLDKREKQGLPROVPESS			
	DEADGNSYQDDDDLRKRSIDTDEPGVAGSIFPDLKSLPALLPLNLDRIPEEEL			
	IDRDNDQKSRNREKELFALHSPDEEPIERLSVPIPKHFGQRLVLCISLKEPI			
	ETEPFASIALYDEVKKKISENEYFDLNSOMGLRPHVPAITITLARSALTSTL			
	YPSDQVFLIVLEKVLQDGGCEGCAEPIRSEMDATRNKLEKLSQADQFCQRIAG			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	TSGDACNLTEFRPRTLTVTNPFKQEGDRSLDEDLKFLADMRKRSYLRRLRITQ			
	LKIDISPAENPHYCLPELLOVKLYPDSRYRPTRELLFPARVYVNTYRNLLIT			
	YPSLTFANRQGSANITVKKVQAGDEPSAKIETKSSCSFSEKAYTAAYVHNHR			
	SPDFHEIKVKLPLATLIDHNLHFTFYVNSCOQKNPLPEPVGTWYIPLMLONRK			
	GFQFVSLERKPOAYSVLSEPEVLPGRKAVDNHKGYNVEVAWSIHQDDPYLDFE			
	FALVALDEHLEPVYIGDMRIEMNNLEIKSISALNSDLEPVERLHLDLKLIL			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	HTTVPLEAVDPVDLDEYLTHPLANDSGPLRLTEPPDDIEVYVSPDRCRTLSA			
	VPESEMDPHVDCIRSTEDMAIVIRKIRKLGTGPNPNTLDKREKQGLPROVPESS			
	DEADGNSYQDDDDLRKRSIDTDEPGVAGSIFPDLKSLPALLPLNLDRIPEEEL			
	IDRDNDQKSRNREKELFALHSPDEEPIERLSVPIPKHFGQRLVLCISLKEPI			
	ETEPFASIALYDEVKKKISENEYFDLNSOMGLRPHVPAITITLARSALTSTL			
	YPSDQVFLIVLEKVLQDGGCEGCAEPIRSEMDATRNKLEKLSQADQFCQRIAG			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	TSGDACNLTEFRPRTLTVTNPFKQEGDRSLDEDLKFLADMRKRSYLRRLRITQ			
	LKIDISPAENPHYCLPELLOVKLYPDSRYRPTRELLFPARVYVNTYRNLLIT			
	YPSLTFANRQGSANITVKKVQAGDEPSAKIETKSSCSFSEKAYTAAYVHNHR			
	SPDFHEIKVKLPLATLIDHNLHFTFYVNSCOQKNPLPEPVGTWYIPLMLONRK			
	GFQFVSLERKPOAYSVLSEPEVLPGRKAVDNHKGYNVEVAWSIHQDDPYLDFE			
	FALVALDEHLEPVYIGDMRIEMNNLEIKSISALNSDLEPVERLHLDLKLIL			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	HTTVPLEAVDPVDLDEYLTHPLANDSGPLRLTEPPDDIEVYVSPDRCRTLSA			
	VPESEMDPHVDCIRSTEDMAIVIRKIRKLGTGPNPNTLDKREKQGLPROVPESS			
	DEADGNSYQDDDDLRKRSIDTDEPGVAGSIFPDLKSLPALLPLNLDRIPEEEL			
	IDRDNDQKSRNREKELFALHSPDEEPIERLSVPIPKHFGQRLVLCISLKEPI			
	ETEPFASIALYDEVKKKISENEYFDLNSOMGLRPHVPAITITLARSALTSTL			
	YPSDQVFLIVLEKVLQDGGCEGCAEPIRSEMDATRNKLEKLSQADQFCQRIAG			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	TSGDACNLTEFRPRTLTVTNPFKQEGDRSLDEDLKFLADMRKRSYLRRLRITQ			
	LKIDISPAENPHYCLPELLOVKLYPDSRYRPTRELLFPARVYVNTYRNLLIT			
	YPSLTFANRQGSANITVKKVQAGDEPSAKIETKSSCSFSEKAYTAAYVHNHR			
	SPDFHEIKVKLPLATLIDHNLHFTFYVNSCOQKNPLPEPVGTWYIPLMLONRK			
	GFQFVSLERKPOAYSVLSEPEVLPGRKAVDNHKGYNVEVAWSIHQDDPYLDFE			
	FALVALDEHLEPVYIGDMRIEMNNLEIKSISALNSDLEPVERLHLDLKLIL			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	HTTVPLEAVDPVDLDEYLTHPLANDSGPLRLTEPPDDIEVYVSPDRCRTLSA			
	VPESEMDPHVDCIRSTEDMAIVIRKIRKLGTGPNPNTLDKREKQGLPROVPESS			
	DEADGNSYQDDDDLRKRSIDTDEPGVAGSIFPDLKSLPALLPLNLDRIPEEEL			
	IDRDNDQKSRNREKELFALHSPDEEPIERLSVPIPKHFGQRLVLCISLKEPI			
	ETEPFASIALYDEVKKKISENEYFDLNSOMGLRPHVPAITITLARSALTSTL			
	YPSDQVFLIVLEKVLQDGGCEGCAEPIRSEMDATRNKLEKLSQADQFCQRIAG			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	TSGDACNLTEFRPRTLTVTNPFKQEGDRSLDEDLKFLADMRKRSYLRRLRITQ			
	LKIDISPAENPHYCLPELLOVKLYPDSRYRPTRELLFPARVYVNTYRNLLIT			
	YPSLTFANRQGSANITVKKVQAGDEPSAKIETKSSCSFSEKAYTAAYVHNHR			
	SPDFHEIKVKLPLATLIDHNLHFTFYVNSCOQKNPLPEPVGTWYIPLMLONRK			
	GFQFVSLERKPOAYSVLSEPEVLPGRKAVDNHKGYNVEVAWSIHQDDPYLDFE			
	FALVALDEHLEPVYIGDMRIEMNNLEIKSISALNSDLEPVERLHLDLKLIL			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	HTTVPLEAVDPVDLDEYLTHPLANDSGPLRLTEPPDDIEVYVSPDRCRTLSA			
	VPESEMDPHVDCIRSTEDMAIVIRKIRKLGTGPNPNTLDKREKQGLPROVPESS			
	DEADGNSYQDDDDLRKRSIDTDEPGVAGSIFPDLKSLPALLPLNLDRIPEEEL			
	IDRDNDQKSRNREKELFALHSPDEEPIERLSVPIPKHFGQRLVLCISLKEPI			
	ETEPFASIALYDEVKKKISENEYFDLNSOMGLRPHVPAITITLARSALTSTL			
	YPSDQVFLIVLEKVLQDGGCEGCAEPIRSEMDATRNKLEKLSQADQFCQRIAG			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	TSGDACNLTEFRPRTLTVTNPFKQEGDRSLDEDLKFLADMRKRSYLRRLRITQ			
	LKIDISPAENPHYCLPELLOVKLYPDSRYRPTRELLFPARVYVNTYRNLLIT			
	YPSLTFANRQGSANITVKKVQAGDEPSAKIETKSSCSFSEKAYTAAYVHNHR			
	SPDFHEIKVKLPLATLIDHNLHFTFYVNSCOQKNPLPEPVGTWYIPLMLONRK			
	GFQFVSLERKPOAYSVLSEPEVLPGRKAVDNHKGYNVEVAWSIHQDDPYLDFE			
	FALVALDEHLEPVYIGDMRIEMNNLEIKSISALNSDLEPVERLHLDLKLIL			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	HTTVPLEAVDPVDLDEYLTHPLANDSGPLRLTEPPDDIEVYVSPDRCRTLSA			
	VPESEMDPHVDCIRSTEDMAIVIRKIRKLGTGPNPNTLDKREKQGLPROVPESS			
	DEADGNSYQDDDDLRKRSIDTDEPGVAGSIFPDLKSLPALLPLNLDRIPEEEL			
	IDRDNDQKSRNREKELFALHSPDEEPIERLSVPIPKHFGQRLVLCISLKEPI			
	ETEPFASIALYDEVKKKISENEYFDLNSOMGLRPHVPAITITLARSALTSTL			
	YPSDQVFLIVLEKVLQDGGCEGCAEPIRSEMDATRNKLEKLSQADQFCQRIAG			
	KYRPFAMTALHININIVSAGSLERDSREVEIISNGERKGSSEERNISIVGRSLERT			
	TSGDACNLTEFRPRTLTVTNPFKQEGDRSLDEDLKFLADMRKRSYLRRLRITQ			
	LKIDISPAENPHYCLPELLOVKLYPDSRYRPTRELLFPARVYVNTYRNLLIT			
	YPSLTFANRQGSANITVKKVQ			

BASE COUNT	2082	a	1384	c	1432	g	1930	t
ORIGIN								
Query Match	33.5%	Score 2132.8;		DB 6;	Length 6828;			
Best Local Similarity	60.4%	Pred. No. 0;		Mismatches 2262;		Indels 273;	Gaps 10;	
Matches 3863;	Conservative	0;						
OY	23	ccgagcgccgcgccttcgcgcacagaatcaagagcaggtgycgcgcagagtgycggaagc	82					
Db	26	ccgagcgccgcgccttcgcgcacagaatcaagagcaggtgycgcgcagagtgycggaagc	85					
OY	83	agggtgcccggaagcagtggtctcccccacatccagcagcgctgcgcgcagctc-----	137					
Db	86	agatctccgcgacaaatatagtggtcttccccaactgcctcaaaaaccttatattgttgcca	145					
OY	138	-----ccctgggggtcccaactgactacgaaatttcgcagccctcgagacttgag	184					
Db	146	atatatcccatcacacacagtcgcccttacggaagcagatatactcaatgagatttggaag	205					
OY	185	atgtaattctcgagccgcgcacagatgctgagcccgccctcaaggaaactggtagaat	244					
Db	206	atttacctctatctactccttggcctgctgagattctgggcctttacgggatttgattgaat	265					
OY	245	tccagctgaactgagggcgtcgtcgcagcccggaagtcgcgcgacacagcgccgc	304					
Db	266	tttccttcagatgatttgaaagttgattatagtcctcgagactcgcaaacctctgtttcag	325					
OY	305	ggatccccaaggaatgaanaactgtagtcgccagtgagggcgcggtgtagatglatatg	364					
Db	326	ctgtacctgaagaaatgaaatgagatccacatggtttagagactgtatmaaacgtttatcacg	385					
OY	365	aggactgggtcaattgtccacagaagtatcagtaactgagtgaagatcagcccgta	424					
Db	386	aagactgggcgaattgtcatcagaaattatcatpaattgggaacacgagtttaattcccaata	445					
OY	425	ccaacacacacgcggygagcgaacagaaggtcctcccccgcagtgctttgacgaatg	484					
Db	446	catatgatttaacgaagaagaagcaaaaaggtttgccaaaacaagttttgaaatcgatg	505					
OY	485	cttcctgagacgaaggttcgcgcctcgaggaactgaaatgactcccggtgtgctcggtc	544					
Db	506	aagcttccacatggcacaacacttaccacagatgattcamaatgactttaaaacgttcaatgt	565					
OY	545	ccccgaagaacccctcgaaagagtggtgcctctgacatcttcgacccggaagaaactcg	604					
Db	566	caatpaatgatraccccaagggtagctgggacctgtgtagtcttttgacttgaataattcac	625					
OY	605	cagctgactcaatgctgcctctctgctagagcggygcgcgcccaagaagatgttgacggcg	664					
Db	626	tttcctgatgttgcttcccaatttacttgatgacactccaatgatgaagaatagacccgtc	685					
OY	665	gcaatgaaaccttcgagcgacgagccgccccgcgcctgtcacccctctacccggcac	724					
Db	686	agaaatgatgaccaaagaaatacaaacccgtacaaaabaactttttgcttttgacatccatcac	745					

Qy	725	ctgaagagatgaagccgctggaaacgctgttagccgcccagagacaccccgagacatttg	784
Db	746	CAGATGAGGAAAGAACCAATAGAAACGGCTTAGTGTTCCTGATATATCCAAAGAAACATTTTG	805
Qy	785	gacaagaagatctgctgcgaagtgctgtgctcaagttcgaagttgaaattgagccatct	844
Db	806	GTCMAAGACTTCTTGTAAATATGCTTATACCTCAAGCTTGAAATATGAAATCCCAATT	865
Qy	845	ttggagctcttgctctgtatgatatgctggagaaagaatctccgagaaactctact	904
Db	866	TTGCAAGTTTGGCTTATATATGATGTCAGAAAGAAAGAAATTTTCACAAAACCTTTTAT	925
Qy	905	tcgactctgaactcgcgactccaatgaagggctgctccgggtctcatatggaccaacctgcca	964
Db	926	TTGACTTATATCTTGAGCAGATGAAAGGGTTGTATGTCACATATGACCTGCTGCCA	985
Qy	965	tctccacccttgccacctctgcactctctctgtacttaacccctcacctgaactcttc	1024
Db	986	TTACATACCTGGCAAGATAGCAATTTTTCATCATTATTCCTTCCAAAGATGTTTTTC	1045
Qy	1025	tgtctatcaagtctggagaaagtctcttcagcaaggagacatacagttgctgtgagcct	1084
Db	1046	TTGTATATAAGCTAGAAAAAGTCTTACAGCAAGACATTTGAGAGCTGTCAAGAACAT	1105
Qy	1085	acatgtgtctgaaagaatgtgacacagccaagaagaagaagctaaagaagcttgccc	1144
Db	1106	ATATGATTTTCAAGACAGACAGATGCCACCAAGAAATGGAAGAAACTGAGAACTGAGA	1165
Qy	1145	tggcgagccgagcagttctgcaccgcgctgggcccgtacacgatatccctctgcgtgagcg	1204
Db	1166	GTCMAAGCAATAGTGTTCGCCMAAGACTTGGGAAATATCGCATGCTTTGCTTGAGACTG	1225
Qy	1205	ccgtgcactctggccaacatcgtgagcagcgtctggcagctctgacccggagcct-----	1264
Db	1226	CAATCATTTTAAATGAAATATGTTTACAGTGTGGGAGTTTGGAAGAAATTTTCACAGAG	1285
Qy	1258	-----gactcggagggcgagcgccggccgacgcttgacagacgcccgc-----	1300
Db	1286	TAGAAATCAGTACTGAGAGAGCAAAAGGCTGTGGTCAAGAGAGCAATTTCTAGTATG	1345
Qy	1301	---gtccggggccccagaacccgggcccagtagtgggagcagcgcctgaacttctctgct	1357
Db	1346	TTGGCAGACGATCACTTAAAGAGAACACAGTGGAGATGATGCTTTGATACGACGACT	1405
Qy	1358	tccgtccagccaagctaaactgtlcaaaactctcttaagcaagaagccgagcgaactcagt	1417
Db	1406	TTTCGACAGCTACTCTCAGACTGACAAATTTTAAAGCAAGAGAGACCGCTTAACTG	1465
Qy	1418	acgaagacccttcaagttctctgctgacatgagcgccgctgtctccctctgcgagac	1477
Db	1466	ATGAAGATCTCTACAAATCTCTGCTGATATGAGAGGCCATCTTGTCTTACGGCGAC	1525
Qy	1478	tacgtccgtgtgacgtcccaagctaaatgacatttctccgctccctgaaatccccact	1537
Db	1526	TAAAGCACTTATACAGCTCAGCTCAAGATAGACATTTCTCCGGCACCTGAAATATCCCAT	1585
Qy	1538	tctgctctccctctgagctgtctcataatcaagccctaccccggaaccccgaggcccgagcca	1597
Db	1586	ATTGCTTAATCTCGGAGCTGCTTCAAGTGAAGCTTTACCTGACAGTATGAGTTAGACTA	1645
Qy	1598	ccaagagaaattcttgaggttcccgcccgcgaaagtctatgcgccccaatacagcttaagaa	1657
Db	1646	CCAGAGAAATTTGAGAGTTTCCCGCAAGGGATGTTATGTTCCAAACACTACTTACAGAA	1705
Qy	1658	acgtgctgtagctgtacccgacagacgtcctaacttcagcagccgcgaggtctcgtgtgca	1717
Db	1706	ATCTTCTTACATATATACCTTCAGAGCTTATATTTGCCAATTCGTCAAGATTTCTGTAGAA	1765
Qy	1718	acctgtcgtgtgcagctgcagtlacatgaaggcgaagccccagccaggtctctgcgttca	1777
Db	1766	ATATTAACAGTGAAGTCCAGTTTATATGATGAGAGGATTCACACCAATTCGATCCCGTAA	1825
Qy	1778	tcttttcaagctccagctcgtgagaaatttaccggcgaagcccttcaacgggtgtctacc	1837
Db	1826	TCTTTGGTAAATCTACTGCTGTTCAGAAATTTTCAAAAGAACCTTATACGCCGTAGTATTC	1886
Qy	1838	atacaagatccccgagttctacgagagttcaagctgtacatctccagctctgagaaag	1897
Db	1886	ATTAACAGGTCCTGATTTTTCATGAGAAATACAGGTATTAAGCTCTCTGCTACTTTAACG	1945
Qy	1898	agaacatacactgctgttcaactcttaccatgtacgtgcgaagccccggccgggacatg	1957
Db	1946	ACCATCATCTTCTGCTTTTACTTTTATCATGTTATGTTGTCAACAAAACAAATATCTC	2005
Qy	1958	cccttggagacacccgtgggacttaacttgatcccaactgtctgcagcagcgccctgagaa	2017
Db	2006	CTCTTGAACACACGATGGATATACATGGAATACCAATGCTTACGAATGAGCGTTGAAAG	2065
Qy	2018	ccggccctctgtctcccaagtgctgtgacccagcccgcccaactatccgtgtcca	2077
Db	2066	CTGGCAGTTTGTCTTGGCAGTCTCATTTGGAAGAAACACACAGGCTTATCTGTACTGT	2125
Qy	2078	caccgatgtgagcgtctccgggcatgcgtctgggtgagcgtgtacaaaggcgtgttcaagt	2137
Db	2126	CTCTGAGGTTCTCTACCTGGCAGTAAGTAATGGGTAGATATACAAAGCTGTTTTAATG	2185
Qy	2138	tgaagctcaagccgtgtctctctgtgcacccccaggaaccccttaccatgaattctca	2197
Db	2186	TTGAAGTTGTGCTGTTTGTGTATCCATACAAAGAAATCCTTATCTTACAAATTTTGTG	2245
Qy	2198	cccttggtgacgtctcctggagagaggaccccttccactccggctcaagaacatgtgtcga	2257
Db	2246	CTCTGTCATATCTCTGATATGAACACACTGTTCCAGTCCGAATTTGGGACATGCGAATCA	2305
Qy	2258	gcgagggcaacgttgagcaagaagctgtgggcccagttctgcagcactgcgtccgtgcccgc	2317
Db	2306	TGGAATAATTAATTTAGAAATGAAATGGAAGAGCAGTATTTACAGCACTGAATTCACGAC	2365
Qy	2318	ccgaacccctgtgacctctccccaacgaagtgagcaaaagctcgtgtgtgtgtcatca	2377
Db	2366	TGGAACCACTGCTCCGATTTCTTCACTCTCTGCTGATTAATCTGATCTTTTAATTTATA	2425
Qy	2378	ggccccgaatcaatcagctgagcagatgtgaaaccttggccgctggagaccttgaagcaatg	2437
Db	2426	GACCTCCCTGATCTGCTGGCCAAATATGTAACCTAGTGCATGCAATCTTTGACACCAATG	2485
Qy	2438	cccatgtgtacgcttgttcaacggagccttggagagcagcccgagatgcccgggtcaact	2497
Db	2486	CATCAATTAATTAATGACTTCAAAAACCTTGGAAAGAAATCATGACCAGCATGGCGAA	2545
Qy	2498	gcccaagctgtgctcagctgtccactacgcttctgcgctccctgcgagcactgaagccagcc	2557
Db	2546	ACAGGCTTCTGCATCATATATATCATTTATGTTTCCGCTACCAAAATATCTTACCCTAAT	2605
Qy	2558	tcccggaa-----tggggccctccagttgaagttgacgtgcacacttggccgtg	2608
Db	2606	CATCATACACAGGTCCTGAGGGGTTTGGAGAGATCAGTCAATTAATGCAATATGCTATGAT	2665
Qy	2609	gctctgtgcgcccgaagccttcaaccggtgttccaagatataagcagaac	2668
Db	2666	CTGCGGTGAGACTGTCAAGCCTTAATTTAAATGCTTTCGAAGCCTTAGTAATATAGCAATC	2725
Qy	2669	ctgaactcgcgtgtgccccctgtgctgtgatagaagagtt-----	2709
Db	2726	CAGATATATCTGGGAGCTCCACAGTCCACAGATGATGAAGTTTCATCAATATCGGGAGTA	2785
Qy	2710	-----	2709
Db	2786	AGGGTTTAGATCGCTCCAAATCTGTGGTTAAACACTGTGTGTCAAAAGCTGCCCATGG	2845
Qy	2710	-----	2709
Db	2846	GATCAACCCCAAGTCCAAAGTCCAGATTAACAACACAGGCTATGATCGAAGTTGTATTCGTA	2905
Qy	2710	-----tcccgcactctgaacca	2725



QY	4865	atgacacatggtgacagcgccgcgcctctgtagtaccctcgccctgcgcgcgcacgc	4924
Db	5117	atggtctactccacccacacacactcttgctgataatttgacacatcttgagacccga	5176
QY	4925	gcaccctgcgcgctgggcctgctcttcctccagaaactctatccaaacygtctaaagagt	4964
Db	5177	aaatctcttcttgggatgtgttaacatttcagaaattttcattcaattatgttttaagaagaat	5236
QY	4985	ccgcacatctcgcacacacatccctgcgcgcgcacagagagagctctctgcctccggaagact	5044
Db	5237	ctgcggcttcacatgatgtgtgtatctccaatgaagaagatattctctctggaanaattct	5296
QY	5045	tcactgagctgggctcgtgtaggtgtctgcgaacaggcagccggtactctcaacatgggcg	5104
Db	5297	ttactgagtcagacgaattcttgggattactggamaacacactcttctctctatngcgtg	5356
QY	5105	ggctctcagagcggtgatgataggtctcaagaacctctcccatccctcgaaagccacc	5164
Db	5357	gcattctatgaacgaatttatgaattttacaaactacttattctcttatcagtaagctaatc	5416
QY	5165	gtgactacaagaagctctggccgcggtgcacgcgaacatctgaaggctctcaacaagatca	5224
Db	5417	gggatgcacaagaactatccacaaattcatngtmaactttcaaaacacatttcacacaaattg	5476
QY	5225	tgcaccagaatccgcgtctggagacgcggtctcggagcgtattctccgcttgggtctcag	5284
Db	5477	tttcattcaagatgacgtccctggagcggagctgtttggacctattttgcctattttcgctgtggtttatg	5536
QY	5285	ggccaccctctgggtgaccttgatgagcagagctttgtcaaaagagccatcgatcaga	5344
Db	5537	gaaccgaatttggggatgtttggatgaacagaaattgttttacaagagccttcacaaacaa	5596
QY	5345	agcttgcagagatcctcacacgcgcctggagagttctcacaacgagagatcttgcaagacg	5404
Db	5597	aaacttcagagatattctcacagatttgagaggaattttacggaanaaagatttgacagagatg	5656
QY	5405	tcgttgagatatataaagaactcttaacctgtgtagcaagtlccaaagcttgactacacaaag	5464
Db	5657	tggttgaagtatataaaagactcttaacctgtgtagcaaggtmaattagatctcctaacaagc	5716
QY	5465	octacatccagatacagtagtggaaacgcgtactttgatacctaagagctcaagaaccggg	5524
Db	5717	catatatttcagatttacctatngtgagaccattcttgtagcacatgtagatgaagaagacgaa	5776
QY	5525	tgacctacttgacgcgaactatggctctgcacatccctgtcttctgcgcgcgcgtccgc	5584
Db	5777	tcacctattttgcacaaaatttacaaattcttgcgtgattatgtagtactgtagacaccttttact	5836
QY	5585	cggatgggcgcgcacacgcgggagctgtcccgagacacaaagcgttaagaagctgtctcaga	5644
Db	5837	tagattggccgcgcacacggggaaattcatgaacaaattcaaaaaggaagcactttctgacta	5896
QY	5645	ccgacacagccttccctatataagactcgcgatccgtgtgtgcacccgggagagagaagg	5704
Db	5897	cgtctcatgcttttctttattatttaaaacaaaggctcaaatgtcactatnaaagaagatgca	5956
QY	5705	tgctgcagccagtgagagtggtccatctgaagacatgcagaagaagacacgcggaactccct	5764
Db	5957	tctttaacaccaaattgaagttgctatgttagagacatgcacaaaaaagacacagagattggcat	6016
QY	5765	ctgcacacgcgcaggaaccacacagatgctaaagtctcagaatggtctctcagagctctg	5824
Db	6017	tttgcaaacacatccagatgccgcgagaccccaaaatgcttcaatgtgtagtctccagagattctg	6076
QY	5825	tagggtccacacgtggaaccagaggtctcccttggaggtgtgcccaggtgttlltagcagagatcc	5884
Db	6077	taggcacacacagatgaatccaggggacttttggaagtgtgccacagttttctctgtaaatfag	6136
QY	5885	cggagagcccaagccttcctgcgcacatacaaaatgtgcgtctctgttcttaagactctc	5944
Db	6137	ctatgacaccaaagccttcttcaacacatcaataataaacctgcacctctgcttttaaaagatttta	6196

[illegible]



Db 2049 GGGTGAAGGACATAGGAGATATTAAATTGAAGTGAAGCTGTTCTCTGTACACAC 2108  
Oy 2169 ccaggacccttacctgacaaattcttccccgtgtccgctcttgag---gaggagac 2225  
Db 2109 CAGAGACACACACTGTGGAGAAATTCTACCCCTCTGCCACTCCCTGGAAAGCCAGGTGAC 2168  
Oy 2226 ctctccattccggtcaagagacactgtgtctgagcagaggcaacgtgtgagcagagctgcg 2285  
Db 2169 CTTCGCCATCCGGTGTGGATGAGAAATACAGGAGATGGGGCTGGACATGAGTGTGA 2228  
Oy 2286 ggcgaattcttgcaactcgtcgtcgtccagccccgaacccctgttgaccttccacca 2345  
Db 2229 GCTACGATCATCTGCTGAACCTCTCCGCCCTGGAGCGCTCTGCTCTTCTCTGACCT 2288  
Oy 2346 cgtgtcgaagaactcgtgtgtgtgtatcaagcccccgacatcaatggtggcagatgt 2405  
Db 2289 GGTGCTGGACAAACCTCTTCCAGCTGTCCGTGACGCCCATGTGTCATGCTGGCCAGAC 2348  
Oy 2406 gaacctgggcgttgaagccttgaagcaatggcccaatgtatgacgctgttcaacggag 2465  
Db 2349 CAACCTTCTCCCAAGTTTCCCTTGAGTCCGTGGTGGCCATCGCCAAAGTCTGCACACAG 2408  
Oy 2466 cctggagagcagcccgatgtgcccgggtgtacatgcacacagctgtgctcacta 2525  
Db 2409 CAAGGACCTGAGCAAGACACACATGGAGGAAGTCCCTGCTGCTCTACGTGCTACTA 2468  
Oy 2526 cgccttctgccttctctgcaactgtgcccagcctccggatggggccccctcgaagtgc 2585  
Db 2469 CGTCTTCCGCGCTGCCAGAGTGCAAAAGGATGTGCCCAAGTGAAGCGCTCCACGTGCC 2528  
Oy 2586 gcaagctgc-----cacatggccgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2633  
Db 2529 CCTAGACCTTCGAGCTACCAACGATATGGCCGACATCAGTGTGCTGTGATGTCAA 2588  
Oy 2634 cctggcgtgttcaagagcatatagcagcagcaacctgtacctgtgcgtgtgccccgtgtc 2693  
Db 2589 GCTGCTGACGGCCCGGGTGTATGAGCAAGTACCAAGACCTCGGGGAGAACACTTCGC 2648  
Oy 2694 tgtgtatgacgaggttctccgcatctgtgcccagcagctgtcttcagagga----- 2744  
Db 2649 ACAGACGAGGAATGAGAACATCATGTCTCAAAAGATCCGCCATGCCAACTGCAGCCG 2708  
Oy 2745 -----gctgctctcagt 2790  
Db 2709 AATGCTTACTATTTCTCTGCGACATAGTATGCTCAAGTTTCACTCGACGCCCAAGGCC 2768  
Oy 2791 atccctccagcagcct-----ggttcttcttccagctcat 2825  
Db 2769 ACCCAGCAAAAAGCATTTCCATGAGGAGCTTGCCCTTCAAGTGTGTGACACCGGAT 2828  
Oy 2826 ggtgaagatltgtgcgtgtgcaactgtgtgtgcccagcagctatagcaaccccgcaagt 2885  
Db 2829 GGTGAAGACATGGCCGACGATACATGACATGACAAACGGGAGATTTTCGGAGGAC 2888  
Oy 2886 ggcgttcccgagcgtctccctggcagacatcaactgctgtgtgtgtgtgtgtgtgtgt 2945  
Db 2889 TCGTTTCTGTGACCGTTTCAATGATGATGATCACTAATTTTATATGTGTGTGACCTGGGA 2948  
Oy 2946 ggtcatcacccgtgt-----ccacaagatgtgtgagctgtgcccagccctcaagc 2996  
Db 2949 AATTGCACCCCTTTAGTAAACACAGAAAGAAATGAACAGCGGAAAGATGAACAT 3008  
Oy 2997 cagctgtgcttctctcactgagctgttccctgtgtgagccggggtgtgtgtgtgtgtgt 3056  
Db 3009 CAGCTGCGCTTCTCTGTATGACCTTCTCTCCCTCATGATGAGGGGCTTTGTGTTAA 3068  
Oy 3057 cctgttccgggcccactcaacagcaggtgtgccaagcgtctccagtctgtccctcaatccagc 3116  
Db 3059 CCTATCTGACATTTATGTGACCGCTGTCAGCCAGCT-----CAGTTAACTTTC 3119  
Oy 3117 agcctgtgtgacctgtgcagtgatcaaccgcatctgtgcaagcagcagctacgt 3176

Db 3120 AACGTCATTTCCATGAGGCTAGAGTTCCTGAGAATCCTCTGTAGCCATGACATTTACT 3179  
Oy 3177 gacctcaactccctctgtcccccgttcaactcagcctgcgctccctccctctgtgtc 3236  
Db 3180 CAATGTGAACCT-----TTTTTTATGAATGCTGATCTGATCTCAACATCTCTTGTGC 3233  
Oy 3237 ctccacacccctccaggagctcaccttctcagcgaagccccggagcccccaaggtgtgacag 3296  
Db 3234 TTCCATATCTTCCAGAACTAACTAACTCTCTGCTCCAGCTTCCAGGACCAAGAAATCGCC 3293  
Oy 3297 catgttcaactgtgtgaccattccggaacacagcacttctcagctgtgtgtgtgtgtgtgt 3356  
Db 3294 CATGTTGATCTGACTTCCGAGTACCGGACAGACACTTCTCTCAACGGGGCTCTCTAC 3353  
Oy 3357 gtagctgacatgtgccccttcgacactgtgaggtgtggaagggatctctgtgtcacaagaagc 3416  
Db 3354 AGAATGCTGCTCTCCCTGATGCTCGAAGGGGAAGGAATCAGCAAAAGTAAAGGAAGC 3413  
Oy 3417 catcagtgctgtgcacagccctctatgttgccatgacacagcaccgccgtacgcgagc 3476  
Db 3414 TGTGATGCAATTTACAGCCCTGCTTAATTTCTACAGACCTGAGCCACGCTGTCAAAC 3473  
Oy 3477 caatgtgaagctcgtgtgtgcccagctgtacactgtccttctcagatltgcagggatc 3536  
Db 3474 AGAGTGAAGGTCAAAATCGCCGCTTACCTTAATGTTGGCATCATTTTGGATGC 3533  
Oy 3537 ctgtccagcgtgtgacttctgtgtgagggccaggttcagcgggtcaagactgtgctcaat 3596  
Db 3534 TTTCACACAGCTCTGTACTTTTACAGTTGACAGTACTCGCAGATACCGCAC----- 3584  
Oy 3597 gctgtactcagacacagaagggagacatltgcyggtatcatcaacccctctgtgc 3656  
Db 3585 -----CAGTGGCTCGGATTAAGAACAAAGAGAGCGCGGCTTATTAACAGATGTGGC 3638  
Oy 3657 catgtccatgt 3716  
Db 3639 TCTGGCATAGCAGGGAAATTAATTTGAATAAACAAGTGAATGT-----GCTGTC 3692  
Oy 3717 accaagcgtcttcogcagcgt 3776  
Db 3693 TTCTTTCCTTATAGCAGTACAAATCTGAAACGGGACACTATCTGCAACTCTATGTAT 3752  
Oy 3777 gttgt 3836  
Db 3753 CTGCTTCTCTGTGATGTAAATGCTGATCAGACCTCATTAAGAAATGATGTGTCTGA 3812  
Oy 3837 cctgtacatcccccagctgt 3896  
Db 3813 CTGCCATCAACGCGACTCAACAGGATTTTATGATCTTATCTGTGTATGTAT 3872  
Oy 3897 ttagtacaaggggaaaaagccttgaacgcatcaacacagcctcaatltcaaaaaatctct 3956  
Db 3873 TGAATTAAGGAAACAGAGTTCTGTACAAAGTCACTACCAAGTCTCAAGATCAAG 3932  
Oy 3957 ggaatatgaagcgcgtgtgtagaagccattctgtgttaccatgtgagctgtgcaagaat 4016  
Db 3933 GGATGTCAAGGCCCGGCTGGAAGAGCTTGTGCGCTGGGGAAGGGGCCAGAGGGAGAT 3992  
Oy 4017 ggttgggggaagtcgt 4070  
Db 3993 GATGCGCCCGCGGCTCCAGAGAACGACGATTTCCAGGCTTAATTAAGAAATTTAGATG 4052  
Oy 4071 ggcgaagagcgttcaacacatctggaagaaaccccaagcgcgtgtgtgtgtgtgtgtgtgt 4130  
Db 4053 GAAGAAAGACAGACACATTTGGCGGACACTAATGGAAGCTAGATTAAGAAAGAGCCGA 4112  
Oy 4131 aatgtacaagcaggt 4190  
Db 4113 GTTAGATCAAGAAAGCTTGATCAGTGGCAATCTGCTACAGAAACATTTAATATATCT 4172  
Oy 4191 ggaacacactgt 4250  
Db 4173 GGATATGACAGAAACATTATCTCAGGCGAGCTCGGCTCTGAGCTGTAAAGACAGCTGCT 4232









4072 cggagagcgctccacacttgagcaaacctccagcccgctgagcaagccaagatgaa 4131  
1820 AGGAAAGATATGCTCAGTGGCGTCAAAAACAGAGAACTTGAGAAATCAAGACACAG 1879  
4132 atggagacagagcgcttgtagaaggaacctggaaccaggaagcctatgcttg 4191  
1880 ATTGAACACGAAACACATGATGTAAGAACTGCTGCTAAGAAAGAACCTTATCTTTTA 1939  
4192 gacacatgagatcatcctgtagagacggtatgcttccagaccggagagacgcttg 4251  
1940 GATACCTTAGAGATTGTTGTTACAGACCGTTCTGTACAGGAATCCAAAGAGACATCTT 1999  
4252 ggggcaagctgaaggtctgctgtacagccctgggcaagctcccaagtgccctcttg 4311  
2000 GGTGGAGTCTMAAAGTGTACTACACACATGGCTGTAAACCAAGTGCAGTTATCTTA 2059  
4312 cagcatgagcctgagccacagagggcgccctgtgtccaaagtcccggaagctgtctgag 4371  
2060 CAACACTGTTTTCTTACACAGAGAGCCTTGTTCAAAAGTTTCCTGAACTTATTTGAA 2119  
4372 gaagacagagctgtgtgtagcagctgtgctgaagctccctacagacatgtagcagcg 4431  
2120 GAAGAGACAGAGCAGTGTGCTGATTTATGCTCAGGCTTCTCCAGACATGTAGCAGTAGC 2179  
4432 atagacacatccgacagcagccagcgctcgctgtacctgtctcatgtagagaaacttc 4491  
2180 ATCGGTACAAATACGTCACACGCGCACTGCTCCCTTACCTAATTAAGGCAAAAGCTTT 2239  
4492 gagatcgagcaacacttgccgctgtgtagatgcaagtcacacatgctctctgctcctg 4551  
2240 GAAGTTGGGAATACCTTTGCGCAGGTTAAATGCAAGTAAACATGTCACATTCCTCTTG 2299  
4552 gtgggagcagcagaaactcaagtagaagacacctgtagcgttcaactcaaacatccctc 4611  
2300 GTGGGCACATCTCAGAAATTTAATGAAGAAATCTTAAGACGTTCTTAAAGACTAATTTG 2359  
4612 acctatctgaagagacacagggcgctggagacagacacctcgcaagcgacggagac 4671  
2360 ACAATATCTTAAAGAAATGCGAATGAGGAAACAACATTTCTGTCAGGTCACAGAT 2419  
4672 ctgagtccaacctgacatcatcctgaagacagcgtgaagaatgaagaacacagag 4731  
2420 CTGGTTTTCANCTCTCCATATGATTTCTTCTGATACGTGAAATGAAGAAACACAGGAG 2479  
4732 gacctgagatctcatcagacctcatgtagaagaatgcccggggtctacccaggtccacg 4791  
2480 GATCTGAATCTGATGATCTAATGTAAGAAATTTCCAAAGGTTACACAGACCTTCCA 2539  
4792 gacctgagctgacctgtgtgtagaacaatgcccgggagacagcgagacacac 4851  
2540 GATCTGGATTTGACCTGTTGCAAGAACATGGCAGGCAAGCACTCAAGACAAATCAT 2599  
4852 ggcagagccgcagctgcatgtagcagcgcgccctcgctgtgtagtacctgacctg 4911  
2600 GCTGAAGCTGACAGTGTCTAGTCCACTCAGCAGACACTTGTGTGTAATTTGAGCATG 2659  
4912 ctgagagacagcagcagcctggtgtagcgtgttccctccagacatcatccaac 4971  
2660 CTGAGAGACCGGAATATCTTCTCTGTGGATGTGAACATTTAGAAATTTCTAAT 2719  
4972 gtgctagagagctccgcacatccgcagacatccctgtgcccagcgagaggtctctgc 5031  
2720 GTTTTGAAGAAATCTGCGCTGATGATGATGATGATGATGATGATGATGATGATGATG 2779  
5032 tccgggaagacatccatcagacatgtaggtgtaggtgtaggtgtaggtgtaggtgtaggt 5091  
2780 TCTGGAATAATCTTACAGAGTCAAGACTGTGGATTTACTGGAACAAGCAGCTGCTTCC 2839  
5092 ttcaacatgggagcgtctacagagcggtgaatgaggtctcaagaacatccatcccatc 5151  
2840 TTCTCTATAGGCTGGCATGTATGAAGCACTTAATGAAGTTTCAAAAGTACTAATTTCTAT 2899

5152 ctggaagccacagctgactcaagaagctggtccggtgtagcagggcaactctcaagagcc 5211  
2900 CATGAACCTTATCGGATGCAAGAAACATATCCAACTTATGTAACATCTCAAGAAAGA 2959  
5212 ttcaacagatcatgacacagatctccggtgtagcggtgtgtcgagatcttccgc 5271  
2960 TTCAAGCAAAATTTCTATCAGAGTACGTGGAGCGAGTGTGGACCTATTTTCCGT 3019  
5272 gtgggtcttcaagcgccacacttcgtgtacctgtagatgagcagaggtgtgtcaagag 5331  
3020 GTTGCTTTTATGTAACCAAGTGTGGGATTTGGATGAACAAGATTTGTTTCAAGAG 3079  
5332 ccatacgaatcagaagctgtagagatctcaaccgctgtagaggtctcaacagagaga 5391  
3080 CCTGCATTAACCAACTTCCAGAGATATCTCACAGATTGAGAGGATTTTCCGAGAAAGA 3139  
5392 tttagcagagacgtcggtgtagatatacaagaccttaccctgtgagcagagctccagct 5451  
3140 TTTGGAGGATGTGGTGAAGTAAAGACTCTAATCCTGTAGCAAGGTAAATTA 3199  
5452 gactcaaaaagcctacatccagatcacgfatgtggaacgtacttgaatacctacag 5511  
3200 GATCTTAACAAGCATATATTCAGATTACCTATGTGAGGCATCTTGAACATATGAG 3259  
5512 ctcaagacccggtgtagactttagaccgcaactaggtgtcgacatctctgtctgc 5571  
3260 ATGAAGGACAGAAATCCTATTTGCAAAATTAACATCTTCTCGATTCATCTACTGT 3319  
5572 acgcgctcagcgcggtgtagcgcgacagggagcgtgcccggagacacaagcgtcaag 5631  
3320 ACACCTTTACTTAAAGATGCGGTGCGTCCATGGGAACTTCAATGAACATTTAAAGAGAG 3379  
5632 acgctgctcagcagcagcagccctccctccatcaagaactcgcactcggtgtgcac 5691  
3380 ACATTTCTACACTGCTGATGATGCTTCTTATATTAACAAAGGTCATATCTCAT 3439  
5692 cggagagagcgtgtctgtagcaggtgtaggtgtccatcgagagatgtagaagaaga 5751  
3440 AAAGAAAGATCATCTTAACACCAATTAAGTGTCTATTTAGGAGATGCAAGAAAGACA 3499  
5752 cggagagctgaccttggcagcagagacccaacagatgtcaagatgtcaagatgtg 5811  
3500 CAGGAGTTGGCAATTTGCAACACATCAGGATCCCGAGACCCAAATGCTTCAAGTGTGA 3559  
5812 cttaaggtctgtagggccacagctgaaacaggtgtccctgtaggtgtgtccaggtgtt 5871  
3560 CTCCAGAGGATCTGTAGGCACACAGTGAATCAGGGGCTTTGGAAGTTGCCAGGTTT 3619  
5872 ttagcagagatcccggaagaccacagctctccgcatcacacaataatgctgctgc 5931  
3620 CTGCTGAATATCTGTAGTGAACCAAGCTCTTCAACATCATTAATTAACCTGAGCTG 3679  
5932 ttcaagactctgcaagaatgtgaggtgtgcgtcgagaaataaagccctgattgg 5991  
3680 TTTAAAGATTTACTTAAAGGTGTGAAGTGTCTTAAGAAAAATTAAGGCTTAATTTGG 3739  
5992 ccggagcagaagagatcacacgctgtagctgtagcgaactatgctgcgctcgaggagct 6051  
3740 CCGGATCAAAAGAGATTAAGAGGAGACTGAGAGAAATATCATTCGCTTAAAGGGGC 3799  
6052 ctgacagccctgttaccagcgctgtcccgagct 6086  
3800 CTACAGCACTGATCAACAGAAAGATCCCTCAGTT 3834

RESULT 8  
AX172880 4144 bp DNA linear PAT 03-JUL-2001  
LOCUS AX172880  
DEFINITION Sequence 7 from Patent WO0142297.  
ACCESSION AX172880  
VERSION AX172880.1 GI:14597915  
KEYWORDS human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 4144)  
AUTHORS Lu, P., Garman, J.D. and Candia, A.F.  
TITLE Clasp-3, transmembrane protein  
JOURNAL Patent: WO 0142297-A 7 14-JUN-2001;  
Arbor Vita Corporation (US)  
location/Qualifiers  
source 1. .4144  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="preliminary Clasp-3 cDNA"  
CDS  
2. .3610  
/note="unnamed protein product"  
/coding\_start=1  
/protein\_id="CAC43727.1"  
/db\_xref="GI:14597916"  
/translation="RNRNRSLNSNPDISGPNPSPPDEVRSIISGKGLDRSNWMT  
GSPKAPMGSNPSPSAESROAMDRCNSKSHHTPTSSLTQITLGRPLPKRIIHEHLA  
QNVCGSVFRESHLOQAMFEELAYKSKVHRLNDKLEAKRSKPERMDIATLV  
STIASIVSRFOKDEMERLNTSLAFPLNDLLSWDGFVSLIKSCYKOVSKLYS  
LPNPVSLRLRDLFLIKCSHBYVTLNPLPSLTPPASPSPSVSSATOSGSESTV  
QDQKIANMELSVPEROAYLAGLVLTETLAVLIDDAEGLGLHKVIMVHNLSSH  
DSDPRYSDPOIKARVAMLYPLIGIMETVLPOLYFTEETHNORGRPLIATDYES  
GSMISOTVMAIAGTSVPOLTRPGSESLTSTSGRCHTFSESSRSLICLLVYKNA  
DETULCKMFTDLVLOLNRLDLVLYVSCFVYKQKVFEEHNSLITPKSKDMRAKE  
EALISGIGAROMVRSRGOLFSPSGSAFGSOELPRKRMTHHROMREKLDKRAE  
TEBHALIDGNLATEANLITLDLEIVOTVSTESKESLGGVGLKVLHSHACQSAV  
YLOHCFATRALVSKRPFLPEETEQQCDLRLRLRSCSSIGTRPSPASLYLIM  
RONEFIGNNEARFKMVPMSLSLVQNEFEELRLSLAT11LYAEDELLETTE  
PDQVDLVNLMHILSDYVAKMHQEDPEMLIDMLVRAKYOTSPDLRTWLQNMAG  
KHBERNHAEAOCLVHSAALVAEYLSMLEDKYLPGCVTFVQSVSSVLEESVSD  
VSPDEEGISGKYFTESGLVGLLEQAAEFESMAGMAYAVNEVYKVLPIHEARNDK  
KLEITIGKLOEAFSKIVHOSGEMRNFYFVFGKFGDLDEOEYKXEPATIL  
AETSHRLGEFGEVREGEVVEYIKDSNVYDKKLDPNKAYIOTIYVEYPTVYMKR  
ITYEDKNILRRPMCTPTLIDGRAGHKEHDFKRLITLSHAPRYKATVNTVTHF  
EITLPIEVALIEDMOKKTDELAFATHODPADPKMLQMYLQGSVGTTVVQGLEAYQV  
LSEIPSDPFLFRHNLRLCFDFTRCEDALRKNSLIGEVQKEYORELEKILSP"  
BASE COUNT 1252 a 829 c 890 g 1173 t  
ORIGIN  
Query Match 22.7%; Score 1446.6; DB 6; Length 4144;  
Best Local Similarity 64.5%; Pred. No. 46-221;  
Matches 2239; Conservative 0; Mismatches 1199; Indels 33; Gaps 4;

Db 626 TATAAACAAGTGTCTTCAAAAGCTTTACTCATTAACGAATCCAGATGTTGTGTCCTTG 685  
Qy 3133 cgaatgaattaccacccgaatccctgtagcaagcaggaactagtagacccttaaccc 3192  
Db 686 AGGCTGATTTTCTACGAATCATCTGACATAGACCATATGTTACTTAACCTTACC 745  
Qy 3193 tgcgtccctctgcaactcaccgactccgctccctccctctgctccaccacactccag 3252  
Db 746 TGCAGCTTACTTACTTCACCTGATCTCATCCTCTGTTTCTTCTGCAACACTCAG 805  
Qy 3253 agctccacactctccagcacaagcccccagcccaaggagtagcagcatcttgcactagt 3312  
Db 806 AGTTTCTGATTTTCTACGAATGACAGACCAAAAGATTTGCAATATGTTGAATTATCC 865  
Qy 3313 ggaacattccgcagcaggaactcctagctggtctctgctgtagcaggaactgtagctgcc 3372  
Db 866 GTGCTCTCCGCAACAGCATTTATTGGCAGGACTTGTTTACAGACTGGCTGTCAATT 925  
Qy 3373 ctggaacactgaagctaaaggggcatctcctgtgcaagaagggcattagcgtgtgac 3432  
Db 926 TTGACCTTATCTATCTAGAGACTGTTGGATTGCATAGAAAGTATCATATATGGTAC 985  
Qy 3433 agcctgtatgtgccaatgacatgacatgaccccgctacgcgcagggccaactgtgaagctcgt 3492  
Db 986 AATTACTCTCCAGTACGACTACAGACCCGCGTACTCTGACCTCAGATTAAGGCTCGA 1045  
Qy 3493 gtggcgaagcttactctgacactgctcttcgattgacagggatgacttgcacgcctcat 3552  
Db 1046 GTGGCATTGTTGATTTACTTACTCTGATTGTTATTCATGGAACGTGACTCAGCTGAT 1105  
Qy 3553 gactctgtgagggcccaagtgcaagcgttcaagactgtgctcaatgcttactcagaacaca 3612  
Db 1106 GATTTTACAAATCTCAATCAATCAACGAGGAAAC-----CAATTGTATGACCACTGAT 1159  
Qy 3613 gaagggcaagggacattctcggttacatcaacccctctgtgccaatgagcactgtcgt 3672  
Db 1160 GATTATGAAGTATGAGCGGAGTATGATATGATCAAGACCGTTGCCATGCAATGCGAAGG 1219  
Qy 3673 ggcgccctagccctgtgctccggcgcaagcatctccaggggccaacagggctctgc 3732  
Db 1220 ACATCGGTCCCTCACTAATCAACAGGCTGCGACTTCTCTCCTCAGCTCAAGTGGCAGG 1279  
Qy 3733 gcaaggtgtgctctctctgtgtagcaagcaggaactgtgctgtgtgctgtg 3792  
Db 1280 CAACACACTTACTTTTACAGCATATCAATCGAAGCCTTTTGAATGTCATCTTGGGTT 1339  
Qy 3793 ctgaaacaacagagccgcgcctcctgcaagcgtgtggccaactgtgacatcccccag 3852  
Db 1340 CTCAAAATGCAAGATGAACAGATTCTACAGAAAGTGGTTACAGATCTCAGCTTGGAG 1399  
Qy 3853 ctggagactgtgtgactgtgactgtgactctgtgctgtgctgtgactgtgactgtgactgtg 3912  
Db 1400 CTAAACCGGCTATTAATCTGCTTATCTGCTGTGCTGTGCTGCTTGAATTAAGGGGAA 1459  
Qy 3913 aaggtccttgaacgaatcaacagactacatccaataaactctgtgatatgaagcgcg 3972  
Db 1460 AAAGTGTTCGAAGAAATGATGATGCTTGACCTTTAAGAAATCAAAAGACATGAGCAAG 1519  
Qy 3973 ctgaggaagcattctgtgtacatcaggaactgtgcaagaatgtgttcgcgaagtct 4032  
Db 1520 CTTCGAAGAAGCTATTCTTGGAGCATAGTGCACAGCAAGAAATGTACGGCGAAGCCGA 1579  
Qy 4033 g-----agaggagccggttttggaaatccagagaacgtgcgtcg 4071  
Db 1580 GGACAGCTGAGAGAGACCCATCTGGAAGTGGCTTTGGAAGTCAAGAAATTTGAGGTG 1639  
Qy 4072 cggagaagcgttcaacacactggaagaaacttaagcccggtgtggaagaagaagatgaa 4131  
Db 1640 AGGAAGATATGACTCTGCGCTCAAAACACAGAGAAAGCTTGGAATAATGACGACAG 1699  
Qy 4132 atggaacaagagcctctgtgtgaaggaacccgtgcaacaggaagcagactgtagtctg 4191  
Db 1700 ATTGAACACGAACGACTGATGTATGTGAACCTGTGCTACAGAAAGCAAACTTAATCATTTTA 1759

4192 gaacatcgtgagatcatcgtgacagcggtgatgtcttcaagaagcccgagagcgcttg 4251  
1760 gATACATTAGAGATGTTGTTTCAGACCGTTTCTGTACGGAAATCAAGAGAGCACTTCTT 1819  
4252 gggagcagtcgtgaagtgctgtacagcctggcagtgcccaagtgccctctcttg 4311  
1820 GGTGGAGTCTMAAGTGCTACTACACAGCATGGCCCTTAACCAAGTGCAGTTATCTA 1879  
4312 cagcatggtccgtcccaagagggcccttggtccaaagttcccggaagctgtgtcgag 4371  
1880 CAACACTGTTTTCTACACAGAGAGCGCTTGTTCAAAAGTTCCGAACTCTTATTTGAA 1939  
4372 gggagacagagctgtgtgcgacccgtgtgcgagctccctcaagacactgtgacgcgc 4431  
1940 GAAGAGACAGAGCATGTGCTCTATTATACCTCAGGCTTCTCCGACACTGTAGAGTAC 1999  
4432 atcagacacatcgcagcagcagcgccctgcgtgtacactgtcatgagacaaacttc 4491  
2000 ATCGTACATATAGTCACACCCAGTCCCTCTTACTACTATAGAGCAAAACTTT 2059  
4492 gtagatcgccacacttgcctgtgaaagatgcaggtcacatgtctctcgtccctg 4551  
2060 GAATTTGGAAATTACTTCCAGGGTTAAATGACAGTACCAATGTCACTATCTCTTTG 2119  
4552 gtggagacagcgcagaaactcagtgaaagcacctgcagactcactcaaaacatctc 4611  
2120 GTGGGACATCTTAGAATTTTATGAAGAAATCTTAAGACGTTCTCTAAGACTATATG 2179  
4612 acctatgtgaagagacatlggggtcgcggagacacacttcgcagagagatccagagac 4671  
2180 ACATATGCTGAAGAAAGATCTGGAAATGAGGAAACAACTTCTCGATACAGGTCAGAT 2239  
4672 ctatattcaacactgacatgatcctgacgagcagcgtgaagaatgaagaacaacagagag 4731  
2240 CTGGTTTCAATCTCCATATGATTTCTTCTGATACGTGTAAGAAATGAAGAACACCGAG 2299  
4732 gaccctgagatctcactcagacatcgtacagaatctgcggcggtctacagaggtcaccg 4791  
2300 GATCCTGAATATGTTGATGTATCTAATGTACAGAAATGCCAAGGTTACAGACACTCTCCA 2359  
4792 gacctcgtgacactggtgtgacagaacagcgccggaaagacagcgagcttggcaacac 4851  
2360 GATCTGGATGTAACCTGGTTGAGAAACATGGCAGGCAAGCACTCAACAGCAACCAATCAT 2419  
4852 gccgaagcgcccgatgcagatgagtcgacgcccctcgtgtcgtacatccctccctg 4911  
2420 GCTGAAGCTGCACAGTGTCTAGTCCACTCAGCAGCACTTGTCTGAATATTTGACATG 2479  
4912 ctgagagacacagcgacacactgcgcgtgtgctgtcgttccctccagaacatcactcaac 4971  
2480 CTGAGAGACCGGAAATATCTTCCGTGGATGTGTAACATTTCAAAATATTTCACTAAT 2539  
4972 gtgctagaagaagtcgcacatctccgaagacatccctgtgcgcgacagagaggtctctg 5031  
2540 GTTTTAAAGAAATCTGGGTCTCAGATGATGTGTATCTTCAATGAAAGAGTATCTGTC 2599  
5032 tcggaagacactcactgagctgagcgctgtgtaggtgtgtctgaaacagcgacgcgtac 5091  
2600 TCTGGAAGAAATCTTACTGATGTCAGACTGTGTGGATTTACTGGAACAGAGCGTGTCC 2659  
5092 ttcaacatggcggtgctcactcagagcggtgaatgaggtctacaagaacctcatcccatc 5151  
2660 TTCTCTATGGCTGGCATGTATGAAGCAGTTAATGAAGTAAAGTAACTATTTCCATAT 2719  
5152 ctgaaagccacgcgtgactcaagaacgtgcgcggtgacagcaacagcgagagcc 5211  
2720 CATAGAGTAATTCGGGATGCAAGAAACTATCCAAATTAATGTGTAACCTTCAAGAAACA 2779  
5212 ttcaacagatcatgacacagagttccggtcgtggagcggtgtctcggagcgtatctccgc 5271  
2780 TTCAAGCAAAATTTGTCATCAAGATAGTGGCTGGAGCGGAGATGTTTGCACCTATTTTGGT 2839

5272 gtgggtctctcagcgcccaactcgtgtgacctgtgatgagcagagttgtgtacaagagag 5331  
2840 GTTGCTTTTATGCAACCAAGTTCCGGGATTTGGATGCAACAAATTTGTTTCAAGAGAG 2899  
5332 ccactgataccgaagcgtgagagatctcacccggtcgtgagaggttctacaagagaga 5391  
2900 CCTGCAATTAACCAACTTTCAGAGATATCTCACAGATGAGAGGATTTTACGAGAAAGA 2959  
5392 ttgtgcagcagcgtcgtgtgatatatcaaaagactctaccctgtgcaaggtccacgtt 5451  
2960 TTGTGAGAGGATGTGTTGAAGTAAATCAAAAGACTCTAATCTGTAGACAGTGAATTA 3019  
5452 gactcaaaaagcctcatccagatcagatcgtatgtgaacccgtacttgttactcagag 5511  
3020 GATCTTAACCAAGCATATATTCAGATTAACCTATGTGAGGCAATCTTGAACATATGAG 3079  
5512 ctcaagagacgggtgacctactctgtgacgcgaactatggtgtcgcacactcgtcttc 5571  
3080 ATGAAGGACGAATCCCTATTTTCAGCAAAATTAACAATCTTGTGATTCATGCTACTGT 3139  
5572 acgcgttcaacgcgcgagtggtgcgcacacgagagctgcgcgagagcacaacagcgtaag 5631  
3140 ACACCCCTTACTTTAGATGGCCGTGCCATGGGAACTTCATGAACAAATTCAAAAGGAG 3199  
5632 acgctgtcagcaccgcagacacgcttccctcaactcaagactcgcacgtgtgtgcac 5691  
3200 ACCATTCGACTACGTCTCATGCTCTTCTTATATTAACAAAGGTCATATGTCATCAT 3259  
5692 cggagagagacggtgtgtcagcgcagtgaggtggtgcactcagagacatgcagaagaaca 5751  
3260 AAAGAGAGATCATCTTAAACCAATTAAGTGTCTTGAAGGACATGCAAGAAAGACA 3319  
5752 cggagcgtggtccttgcaccgcagcagagaccacagatgtctaagatgtcacagatgtg 5811  
3320 CAGGAGTTGGCATTTGCAACACATCAGGATCCCGACAGCCCAAAATGCTTCACATGTA 3379  
5812 ctccaggtcctgttagagcccaacggtgaaagaggttcccggtgaggtcccgaggtt 5871  
3380 CTCCAGGATCTGTAGGACACAGATGATCAAGGGCTTTTGAAGTTGCCCCAGGTTT 3439  
5872 ttagcagagatcccggaagaccacagctcttcgcgcatacaacaaattgctgcgtcgc 5931  
3440 CTGTCTGAATATCTTAAGAGACCAAGCTTTCAGACATCATTAATTAACCTCGACTCGC 3499  
5932 ttcaagagactctgcagaagaatgtgaggaatgcgtgtgcggaataaagggccgtattgg 5991  
3500 TTTTAAAGATTTTAAAGAGGTGTAAGATGCTTMAAGAAAAATTAAGCTTAATTGGG 3559  
5992 ccggaaccgaagagttacacacggtgagctgtgagcgcactactcgcgcgtcggagagct 6051  
3560 CCGGTTCAAAAGAGTATCAAAAGGAATTTG--GGAAACTATCTTGGCTTAAAGAGGCC 3617  
6052 ctgcagccctcgtcttaccagcgctgtcccaagctgtatgagcaccacccc---accgg 6107  
3618 CTACAGCCTATACATCAAGAAAGTCCCTCAGTATTCACCAAGACAGTATGTGTGCCCTG 3677  
6108 cctcaggaactccttgaacagagcaagttccgaaagcgagactcgtagc 6158  
3678 CCACAGAGATTCCTTCAGTGAATGAGCTTTCGCAAAATGATGCTCTAATC 3728

RESULT 9  
AX172932 4144 bp DNA linear PAT 03-JUN-2001  
LOCUS AX172932 Sequence 59 from Patent WO0142297.  
DEFINITION AX172932  
ACCESSION AX172932  
VERSION AX172932.1 GI:14597917  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 4144)



4252 ggggcsagctgctnaggctgtgtctgtacagcctggcgagtgcccaagtgccctctctg 4311  
1820 ggtggagctgtctaaagctgtactacacacatgacctgttaacaaagctgactttatcta 1879  
4312 cagcatgagctgtgcccacagagggccctgtgtccaagttcccgagagctgtgttcag 4371  
1880 caacactgttttcttcaacagagagccctgtgttcaaaatttcttgaactctttatttga 1939  
4372 gaagacagagctgtgtgtccgacccctgtcgaagctcctcaagacagctgtgacccg 4431  
1940 gaagacagagagagtgctgtctgattttatgcttcagccttctccgacactgtaacagatgac 1999  
4432 atcagacacatccgacagacgacgacgctcgctgtactcgtcaatgacagagaatc 4491  
2000 atcgatcaaatatgctcaacacccacgctcccttacttactaattgagcgaataattt 2059  
4492 gaagatggccacacttcccgctgtgaagatgacatgacatgctctccctccctg 4551  
2060 gaagattgggaatgacttggcaggggttaaaatgcaaggtaccatgctacttctccttg 2119  
4552 gtgggagcagcagagacactcagtgaaagacacctgtcgaactcactcaaacacatc 4611  
2120 gtgggacacatctcagaaatttttaagaagattcttaagacgttcttaagaactatattg 2179  
4612 acctatgctgaagagacatgggctcgcgagacacacacttcgacagacagatccagac 4671  
2180 acatattgctgamaagaaatctggaattgagggamaacaaatttcttctatcaggtccgagat 2239  
4672 ctatgttcaacactgacatgacatcgcagcagacagtgaaagatgaagaaacacagag 4731  
2240 ctggttttcaatctccatattgattttcttctgattctgtgtaaaatgaagaaacacagag 2299  
4732 gacccctgaagctcctac 4791  
2300 gatcctgaatattgtgattgattgattgattgattgattgattgattgattgattgattg 2359  
4792 gacccctgagctgac 4851  
2360 gatcctgagattgac 2419  
4852 gacgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 4911  
2420 gctgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2479  
4912 ctcgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 4971  
2480 ctgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2539  
4972 gtgctgaagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 5031  
2540 gtttttgaagaaatcttgcggtctcagatgattgatttctcagatgaaagagatttctgc 2599  
5032 tccggaagacactcactcactcactcactcactcactcactcactcactcactcactc 5091  
2600 tctggaagaaatcttacttacttacttacttacttacttacttacttacttacttactt 2659  
5092 ttaacatggcgcggtctcactcagcgcggtgaatgaggtctcaagaacactcaccac 5151  
2660 tttcttattgcttgcgcatgttgaaagcagtttaagatttcaaaatgacttatttctatt 2719  
5152 ctggaagccacccgtgactacacagagcgtgcgagcgtgacagcgaacacagcagcgag 5211  
2720 catgaagcctaatcgagatgcaaaagaaatcttccacaatttatgtaaaacttcaagaagca 2779  
5212 ttaacaaagatcactcagcagagttccgctgtggagcgcggtgttcggagcgtatctcgg 5271  
2780 tttcagcaaaattgtttcagagactgctggagcgaggtgtttggacaccttattttccg 2839  
5272 gtggagcttcaagcgccacactcgtgtgacctgtgagagagagatttggatcaagag 5331  
2840 gtttggttttatgaaacaaagttcggggaatttgatgaacaaatgtttgtttacaaagag 2899  
5332 ccatcgatcaagagctgagagatctcacacgcgtgagagaggtctcacagagagag 5391

2900 cctgcaaatmaacaaactgtgcagagatattctcagatattgagggagattttcggagaaga 2959  
5392 tttagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 5451  
2960 tttggagagagatgtgtgttggaagtataaagacacttatttcttgaacaaagtgaaatt 3019  
5452 gactacaagaagcctacatccagatcagatgtgtgaacacttacttgaatacctagag 5511  
3020 gatccatacaagagcatattatcagattatcattatgagagcattttgacacattatgag 3079  
5512 ctcaagagcgggtgtgacctacttgaacgcagacatggtgtgttcgaatctcgttctgc 5571  
3080 atgaagagcaaatcaccattttgacacaaaattatcaatcttctcgaattcattgactgt 3139  
5572 agccgcttcaagcggagatggcgcaacagcgagagctgcccagacacaaacgctlaag 5631  
3140 acaccctttactttatgagtgcccggtcccatggggaacttcatgaaacaaattcaaaagag 3199  
5632 acgctgtcagaacacgacacgacctccctacatcaagactgcacatcggtgtgtccac 5691  
3200 accattctgactgactgactgactgactgactgactgactgactgactgactgactgact 3259  
5692 cggagagagagcaggtgtgtgacgacgtgaggtgtgacatcgagacatgcaagaagaca 5751  
3260 aaagagagatcatattttacacacattgaattgctattgagagacatgcaaaaaagaca 3319  
5752 cggagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 5811  
3320 cagagatttgacattttgcaaacatcagagatcccgacaccccaaaatgctttcagatgcta 3379  
5812 ctcaagagcctgtgagggccacacgctgaaacagagttccctcgagaggtgtgcccaggtt 5871  
3380 ctccacagagatctgtatgagcaccacacacacacacacacacacacacacacacacac 3439  
5872 tttagagagatcccgagagac 3491  
3440 ctgcttgaatttccatgagac 3499  
5932 ttaagagacttctgaagaatgtgagagtgctgtcgagaaataaagccctgattggg 5991  
3500 ttttaagatttttactttaaagaggtgagatgctttaaagaaataaagagcttattggg 3559  
5992 ccgagacagagagatcaccacggtgagctgagagcagcactcgtcgctgcgagagct 6051  
3560 cccggttcaaaagagatgtcaaaaggaattg--ggaaacttatttcttgcctttaaagagcc 3617  
6052 ctgcagccctgtctacacacagcgtcccgacgtgagtcacacacacacacacacacac 6107  
3618 ctacagcccttgaatcagaaagatccctcagttatccaaagcagatgtgtgtccctg 3677  
6108 cctcaggaactccttgaacagagacagtttcogaaagcagacactcgtgagc 6158  
3678 ccacagagatttcttctcagtcagatgagctttcgcaaaatgagatcttcaaac 3728

RESULT 10  
AX172980  
LOCUS AX172980 4200 bp DNA linear PAT 03-JUL-2001  
DEFINITION Sequence 107 from Patent WO0142297.  
ACCESSION AX172980  
VERSION AX172980.1 GI:14597958  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct.  
artificial sequence.  
REFERENCE 1 (bases 1 to 4200)  
AUTHORS Lu, P., Garman, J. D. and Candia, A. F.  
TITLE Clasp-3 transmembrane protein  
JOURNAL Patent; WO 0142297-A 107 14-JUN-2001;  
Arbor Vita Corporation (US)  
FEATURES  
source  
1..4200

```

/organism="synthetic construct"
/db_xref="taxon:32630"
/note="polynucleotide fragment"

BASE COUNT      1262 a      844 c      907 g     1187 t
ORIGIN

```

Query Match	22.7%;	Score 1446.6;	DB 6;	Length 4200;
Best Local Similarity	64.5%;	Pred. No. 4e-221;		
Matches 2239;	Conservative	0;	Mismatches 1199;	Indels 33;
				Gaps 4;

QY	2713	cgatccctggcagcaagctgctctacagagagctgctgagctgagctgagctgagcagcagc	2772
Db	321	CGCTTACCAACTTAAAGCTTTTTCACGAGAGCTGGCTTTCAGTGGGTGTTTGGAG	380
QY	2773	agtcgcctacgcagagccatccctccagaagcctggtctctctccagctcaatgctgaaag	2832
Db	381	GGCAGCGCTTTCGGGAATCACGCTTTCGACAAAGCGCTGTTCTTTTGAATTAAAGTAAG	440
QY	2893	agatagcgctgctgactgctgctctgcccagcgactatagacaccccgcaagctgcgcctc	2892
Db	441	AGCATGGTGCACCATTTATCTTTATGATTAACCTTGAGGCTCCAAAGAAAGTCCTTTT	500
QY	2893	cccgagcgccttcctggagacatacactgcctctgctggagctctctgaggccctggagagccac	2952
Db	501	CCAGAAAGTTTCAATGAGATACATTCGACACTCTTGTCACACGATTCCTAGTGAATAGTT	560
QY	2953	accgcgtcccaagaatgtgagctggccgagaccctcaagcgcagctgcgtctcttc	3012
Db	561	TCACGATTTTCAGAGGACACAGAAATGGTTGAGAGACTCAATACAAACCTTGACTTCTT	620
QY	3013	ctcagagaccctctgcccctggtagccggggcttctctcagccgcggtccggagccac	3072
Db	621	CTCAATGATCTGTGTCTCTTTATGACACAGAGATTTGTTTACCTTATAAAGCTCTGC	680
QY	3073	tacaagcaggtgagccacgcgcgtccagctcagctcagctccctacacgaagccctgcagccctg	3132
Db	681	TATAAACAGGTGCTTCAAAGCTTACTCATCTTACCGAATCCAGTGTCTGAGTCTCTTG	740
QY	3133	cgcattgaaatcccccgaactctctgtgagccagcagagcaactagtgcacctcaacctccc	3192
Db	741	AGCGTGGATTTTCTACGAATCATCTGCAGTATGAGACATATGTTACATTAATTAATCC	800
QY	3193	tgcctgccccctgctacactccagcctccgccccctccctctgtccctccacacacctccag	3252
Db	801	TGCAGCTTACTTAATCTGCACCTGCATCTGCATCACCCTTCTGTTCTTCTGCAACATCTCAG	860
QY	3253	agctccaaccttccagcccaagcccccggaagccccaagtgtagccagcagctgcgaactgagt	3312
Db	861	AGTTTCGGATTTTCTACGAATGATACAAACCCMAAAGATTCGAATATGTTTGAAATATATCC	920
QY	3313	gagacattccggcagcagcaactctcctagctggtgctctctgctagagagctgagcagcgcc	3372
Db	921	GTCGCTTTCCGCCAACACGACTTATTTGGCAGCACTTGTGTTAACAGAGCTGGCTGTGCAAT	980
QY	3373	ctcgaacctgagagcctgaaaggggcatctcctgtgtcacaagaaggccatcaagctgtgcaac	3432
Db	981	TTAGACCCCTGATGTGAAGGACTGTTGGATTGCATTAAGAAATCATCAATATGTGTAAC	1040
QY	3433	agcctgctctgtgggccatgagacactggaaccccgctaagccgagggccacatgtgaaagctcgt	3492
Db	1041	AATTTACTCTCCAGTCACAGACTCAGACCCCGGGTACTCTCAACCTCAGATTAAGGCTCGA	1100
QY	3493	gtggccgagcctctacatgctgcaactcttgatgtgaaaggaatcccttgccaagctgcgat	3552
Db	1101	GTCGCGATGTTTATCTACTCTGTATGTAATGTAATCAATGGAACATCTACTACCTGAT	1160
QY	3553	gaacttgcctgagggcccaaggtcagcagcggtlcaagactggctcccaatgcttctactccagacaa	3612
Db	1161	GATTTATGGAACATCAACATCAACGAGGAAGAC-----CAATTTGTAATACCACTGAT	1214
QY	3613	gaagcgaaaggggaatctgctggtgtacatacaaacctctgtgtgcataggcactgtcgtg	3672

Db	1215	GATTTTGAAAGTGAAGAGCGGAAGTAATGTAAGCCAGACGCTTGCCATGCGCAATTCCGAGGG	1274
OY	3673	ggccccctagagccccctgagctcccgcgcaagatctcccaagggcgccacaagggctctcgc	3732
Db	1275	ACATCGGTCCTCAACCTAACCAAGGCGCTGGCAGTTTCCTCTCCACGTCACAGATGGGACAG	1330
OY	3733	gaagctctgcccctctcgtcgtcgaagtcgaagccggagactctgtgcgtgtgtgtgtgtgt	3792
Db	1335	CAACACACTACTCTTTTCAGCGAAGTAACAAAGTCGAAGGCTTTTGATATGTCCTACTTTGGGTT	1394
OY	3793	ctgaaaaacacccgagccggcgtccctccgaagctggcgccactgagcactcccccag	3853
Db	1395	CTCAAAATAATGCAAGTAAGAAAGCTCTACACAAAGTGTTTACATATCTCCTACTCTTGCGAG	1455
OY	3853	ctgggaagctctgtgtgagctctgtctaccttccctggctgctgcttgaatagaagggaaa	3912
Db	1455	CTAAACCGGCTATTAGATCTGCTTTAATCTCTGTGTCTCTTGTCTTTGAGATGTAAGCGAAA	1514
OY	3913	aagcctcttgagcatcacaagcctcacaattcaaaaatctctgatatgaagacggcg	3972
Db	1515	AAAGTGTGGAACGAATGAATAGCTTACCTTTAAGAAATCAAAAGACATGAGACCAAG	1574
OY	3973	ctagaagaagccattctggtgtacatctgagctcgacaaagaaatggtctcgagaaatcgt	4032
Db	1575	CTTGAGAAAGCTATTCTTGAGAGCATGAGTGGCCAGACCAAAATGATGACGGAGACCGA	1634
OY	4033	g-----agaggaagccgcttggggaatcccgagagaagctgtgcgtg	4071
Db	1635	GGACAGCTCCGAGAGAGACCCCATCTGGAGAGTGCCTTTGGAAAGTCAAGAAATTTGAGGTGG	1694
OY	4072	cggaaagcgtcacacacactgtgaagcaaacctccacagccgctgagcaagaagccaaagatga	4131
Db	1695	AGGAAGCATATGACTCACTGGCGTCAAAACACAGAGAACCTTGACAAATCAAGAGCACAG	1754
OY	4132	atggaacccgagagcctctgtgtggaaggaacactgtgcaaccgagcaagcctagtgtcttg	4191
Db	1755	ATTGACACGGAAGCATGATTTGATGAAACCTGGCTACAGAAAGCAAACTTAATCATTTTAA	1814
OY	4192	gacacactgagatcatalctgtgcagacgctgtatctcttcgaagcccggaagagcgcttg	4251
Db	1815	GATCATATTAGAGATTGTTGTTTACAGCCGTTTCTGTAAAGGAATTCAAAAGAGACATTTCTT	1874
OY	4252	ggggcagctgctgaaaggtgtgtgtgtacagcctgtggcagatgtcccagaatgtccctctctg	4311
Db	1875	GGTGGAGGCTTAAAGAGTGCTACTACACAGATGCGCTGTAAACCAAAAGTCACTTAATCTA	1934
OY	4312	cagaatgctcgtgcccacccgaagggccctgtgtccaaattcccgagatcgtctgttcgag	4371
Db	1935	CAACACGTGTTTGTCTACACAGAGAGCGCTTGTTTCAAAATTTCTCGAAGCTCTTAATTGAA	1994
OY	4372	gaagaaacgaagctgtgtgtccgaactgtgcctgaagctcctaagaaactgttggacgcgc	4431
Db	1995	GAAAGACAGAGCAGTGCTGTGATTTATGCTTCAGAGCTTCTCCGACACTGTAGCAGTAGC	2054
OY	4432	atcagacacatccgacagcgcgcgcgcgcctgtcttacctctcaatccgacgaagaaatc	4491
Db	2055	ATTCGTTCAATACGCTACACCCCAAGTGCCTCCCTTTACTTAATATAGCGCAAAACTTT	2114
OY	4492	gagatcgccacaaactcttgccgtgtgaaagatgacagtgatcatatctctctgtccctg	4551
Db	2115	GAGATTGGGAATTAATCTTGCCAGAGGTTAAATATGCAAGGTATCCAAATCTATCTCTCTG	2174
OY	4552	gttggggaacgacgagaactctcagtgaagaagcaactgtcgacgttcaatcaaaacatccctc	4611
Db	2175	GTTGGGCAACATCTCAGAAATTTTATGAAAGAAATCTTAAAGACGTTCTCTAAACATATATTG	2234
OY	4612	accatgctgaagaaagacatggggcgtgcyggagacgacactctcgagagcaagctcaagac	4671
Db	2235	ACATATCTCTAABAAGATCTCGAATGTAGAGGAAACAAACATTTTCTCGATCAGGTCCAGGAT	2294
OY	4672	ctgaatgtcaactgacatgacatcctgacgagacacggtgaagaatgaagaaacacagag	4731
Db	2295	CTGGTTTCAATCTCATATGATATCTTCTGTACTGTGAATAATGAAGAACACACGAG	2354

QY 4732 gaacctgagatctcatcgaactcatatgtaagaattgcccggggtaccagggtctcaacg 4791  
 DB 2355 GATCCCTGAATATGTGATTTATCTAATGTACAGATTGCCAAGGTTACCAAGCATCTCTCA 2414  
 QY 4792 gaacctcggtcgtggttggaagaacatggtcggaagaacggtgagatgtggcaacac 4851  
 DB 2415 GATCTCGATTGACCTGGTTGGCAAGACATGGCAGAACCTCAGAACGAAGCATATCAT 2474  
 QY 4852 gccgaagccgcccagtgcatggtgcaagcgccgcccctgtgtgctgagtaacctgcacctg 4911  
 DB 2475 GCTGAAGCTGACAGAGTGTCTAGTCCATCAGCAGCATCTGTTGTAAGATATTGAGCAT 2534  
 QY 4912 ctgtagagaccagcgccactgtccctgtgggtcgtgttccctccagaacatctcatccac 4971  
 DB 2535 CTGGAGAGCCGGGAATATCTTCTGTGGGATGTGAACATTTCAAAATTTTTCATCTAAT 2594  
 QY 4972 gtgctgagagagtcgcgcacatcccgacatccgtgtgcccgaagaagggtctctgcg 5031  
 DB 2595 GTTTTGAAGAAATCTCGGCTCTCAGATGATGTGATCTCCAGATGAGAGGATATCTGC 2654  
 QY 5032 tccgggaagcaactcaatgagctggtggtgaggtgtgctggaacaggaagcggtctac 5091  
 DB 2655 TCTGGAATTAATCTTACTGATCAGGACTGTGGGATTTACTGGAACAGCAGCTGCTTCC 2714  
 QY 5092 ttacacatggggcggtctctacagagcggtgaatgaggtctacaagaacctcatcccatc 5151  
 DB 2715 TTCTCTATGCTGGCTGCGATGATGAAGCATTAATGAAGTTTAAAGATTAATTTCTAAT 2774  
 QY 5152 ctggaagccacacgctgactacacaagaagctgtgcgcggtgcaagcaaatgtagaagacc 5211  
 DB 2775 CATGAAGCTAATCGGATGCCAAGAAACATATCACAATTCATGATTAACCTTCAAGAGCA 2834  
 QY 5212 ttacaagaatcatgacacagagatccggctgtggagcggtgtctgggaagctatccgcg 5271  
 DB 2835 TTTCAGCAAAATTTGTCATCAGATGACTGCTGGAGCGGATTTGGCAGCTAATTTTGGT 2894  
 QY 5272 gtggagcttctacggcgccactgtgagctgtgagctgtagaagattgtgtacaaggag 5331  
 DB 2895 GTTGGTTTATGGAACCAAGTTGGGATTTGGAGTGAACAAAGAAATTTGTTTCAAGAGG 2954  
 QY 5332 ccatacatcagaagctgtgacagatctcacacggctgtggaagattcttacaagagaga 5391  
 DB 2955 CCGCAATTAACCAAACTTGCAGAGATCTCACAGATTGGAAGGATTTTACGGAAGAAGA 3014  
 QY 5392 ttgtgagagagatggtgtgagattcaagaagctcttaccctgtgtgacaagtccaagct 5451  
 DB 3015 TTGGAAGAGATGTGTTGAAGTAACTAAAGACTCTAATCCTGTAGACAAAGTGAATTA 3074  
 QY 5452 gactcaaaaaggcctacatccagatcaagatgtgtggaacggttacttggatacccaagag 5511  
 DB 3075 GATCCTAACAAAGCATATATTCAGATTACCTATGAGGCCATTAATTTGACACATATGAG 3134  
 QY 5512 ctcaagagacgggtgagctacttaccgcaactatggtggttcgcacattccctgtctgc 5571  
 DB 3135 AAGAAGGACAGATCACTATTTCCAGCAAAATTTACAAATCTTCGTCGATTCATGTACTGT 3194  
 QY 5572 acgctgttcaagcggtgagcggtgacacaggggagcggtgtcccgagcaacaacagcgtlaag 5631  
 DB 3195 ACACCTTACTTATTAGATGGCGTCCCATGGGAACTTCATGAACTTCAAAAGGAAG 3254  
 QY 5632 agctcgtctagacacgacacgcttccctcatcaagaagactcgcatccgtgtgtgcac 5691  
 DB 3255 ACCATTCTGACTAGCTCTATGCTATGCTTTCTTAATTAATAAACAGGCTCAATGTCTACTAT 3314  
 QY 5692 cggaggaagacggtgtgtagcggtgaggtgtggtcctgagacatgcaagaagaagaca 5751  
 DB 3315 AAGAAGAGATATCTTAACACCATTTGAAGTGTCTATTTGAAGAGATCAGCAAAAAGACA 3374  
 QY 5752 cggagagctgtgcttgcacacgagcagaaccaacacatgtctaagatgtctagaatgtg 5811  
 DB 3375 CAGGAGTTGGCTTTTGAACAACATCAGATCCCGCAGACCCCAAAATGCTTCAGATGCTA 3434

QY 5812 cttaaggtctgtagggcccaacggtgaaccaagggtccctggaagtggtgccagtggtt 5871  
 DB 3435 CTCCAGGATCTTAGAGCACACAGATGATCAGGGCCCTTTGGAAGTTGGCCAGGTTT 3494  
 QY 5872 tttagcagagatcccggaagaccaccaagctcttccggtatcaacaatattcggtctgc 5931  
 DB 3495 CTGCTGAATTAATCTAGTACCCAAAGCTCTTCAGACATATATTAATTAACGCGACTCTGC 3554  
 QY 5932 ttcaagactctcgaagaaatgtgagatgcgtgtcggaagaataagagccggtatggg 5991  
 DB 3555 TTTAAGATTTTACTAAAGGTGTGAAGATGCCCTTAAGAAAAAATATAGACTTATATGGG 3614  
 QY 5992 ccggaacgaagaagagatcaacacggtgagctgtgagcgcaactatgcgcgtcggtggaagct 6051  
 DB 3615 CCGGTTCAAAAGAGATATCAAAAGGAATTTG--GGGAACATATCTTCCCTTAAAGAGGCC 3672  
 QY 6052 ctgcaagccctgtctaccacagcgctgcccagtgtagtgcaccacccc-----accgg 6107  
 DB 3673 CTAAAGCCCTTGAATCACAAGAAAGTCCCTCAGTTATCCAAAGCCAGTATTTGCTTGTCTG 3732  
 QY 6108 cctcaaggaactccttgaacagaagaagattccggaaggcgaacctctgagc 6158  
 DB 3733 CCACAGAGATTCCTTCAGTGCAATGAGCTTTGCAAAATGATGATCTTAAC 3783

RESULT 11  
 BC021414  
 LOCUS BC021414 1900 bp mRNA linear ROD 18-JAN-2002  
 DEFINITION Mus musculus, clone IMAGE:5005183, mRNA, partial cds.  
 ACCESSION BC021414  
 VERSION BC021414.1 GI:18204020  
 KEYWORDS  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS Strausberg,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK  
 COMMENT NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: <http://www.nisc.nih.gov/>  
 Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)  
 Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjumin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, O.L., Mastaglio, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantirip, S., Thomas, P.J., Tjongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
 Series: IRK Plate: 39 Row: d Column: 18  
 This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.  
 Location/Qualifiers  
 1. 1900  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5005183"

FEATURES  
 source



BASE COUNT	472 a	512 c	517 g	399 t	ORIGIN
CDS					<p> /tissue_type="Mammary tumor," MMTV-LTR/INT3 model, 5 month  old mouse. Taken by biopsy.  /clone_lib="NCI_CGAP_Man2"  /lab_host="DH10B"  /notes="Vector: pCMV-SPORT6"  &lt;1..1710  /codon_start=1  /product="Unknown (protein for IMAGE:5005183)"  /protein_id="AAH21414.1"  /db_xref="GI:18204021"  /translaction="PRVRSADLYLMKRONFEGHNFARVKKMLVTNSLSLVSGETONTSESE  EHLKSLKTLITYAEVDEIGLDSSTFAAYODLMTNLMHILTDIVKREHODEPMLMD  LMYIARGVYOSGPDRLRTLTKNOMNAGHAEVGNAAEQACWHAALVAEYLTALDESR  HLPGVGSFQVSNVNLSESAISDDILSPDEBEGSCSNFTELVGLLEDAAGFTMM  GGIEAVENEYKKNLPILEAHADRYKLAIAVHGKQOEAFRTIMHOSGMEHFYGFYFV  GGVETREGDLDEDEEFVKKPESTITKLAISHGLEEFYEREGDDVGLITKQSNPDKSR  LDPKAIVQITIVYEPHMDYDYLKDRVYFPRNYSKRAFLCTPTTPPGRAAGELAPDH  KRKTLSTEHAFYPIKTRIRYVCHREETVLPVEVAIEOMKRTKTELPAFQDDPDKR  MLQVNLGVSQPTVNGQPLEVAOVFLEISPDPELFRHNNKLTCLFQDFCKCDALR  KNKLLIGPDDQKEYHRELEHRYSRLREALQPLTLRQLPLAPSSTSLRSSSMNRSSFRRK </p>

[illegible]

QY	5108	ctcagagcggtgtaatgtagtctacaagaactcaatcccatctctgtaagccacggt	5167
Db	662	TGTATGAAGCGGTGAACGAAGTCTTACAAAACCTTATCCCTTGTAAGCCCAACAG	721
QY	5168	actcaaaagcttgcgcggtgtcacgcgaactctgaaagccttcaccaatcatcgc	5227
Db	722	ACTACAAAGAGCTTGGCTTGCCTGCAGCGGAACCTGCAGAGGCTTTACCAAGATTATGC	781
QY	5228	accagagctccggtcttgagacgcggtgtctcggaacgtattccgcttggtctcaagcg	5287
Db	782	ACCAAGAGCTTGCGTGGAGACCTGTATTTGGACATATTTCCGAGTGGGCTTCTATGSCA	841
QY	5288	cccacttcggtgacctgtatgtagcagagttgtgtcaaaagagccatgatcatcaagaagc	5347
Db	842	CACGATTTGGTGAACCTGTGAATGAACAGAGTTTGTGACAAAGAACCGTCAATACAGAAAGC	901
QY	5348	tgcagagatctccacacgcgcctgagagagttctcaacggagagatlttgccagacgtcgc	5407
Db	902	TTGCAAGAGATCTCAACCCGGCTGGAGAGTCTTATACGGAAGGTTCCGGGATGTATGGG	961
QY	5408	ttgagatltcaaaagactcttaccctgttgaaagtcceaaagctttgacttcaaaaagcct	5467
Db	962	TAGAGATCATCAAAAGATTCTTAACCCATGACAAAGTCACAAAGTGGACCCACGAAGGGGT	1021
QY	5468	acatccagatcaagtatgttgaaacggttctataccacagctcagatctcaagaacgcgttga	5527
Db	1022	ACATCAAGATACCTATGTGGAGCGCAATTTGCACACTTATAGCTCAAGATGGGGTGA	1081
QY	5528	ccctacttgaccgcgaactatggtgcttcgcacatctctgtctgcacgcgcgttcaagccgg	5587
Db	1082	CTCTATTTGATCGGAACCTATGAGGCTGGGGGCTCTCTCTTCTGCAACACCTTTCACACAG	1141
QY	5588	atgagcgcgacacaggggagctgcgcgagaaacacaaagcgtaaagcgtctcagacacg	5647
Db	1142	ATGGACGTGGCCACGGAGAGTTGGCCCAACAGCACAAGCAAGAGAGCTGTGAGACAGG	1201
QY	5648	accaagccttccctcatcaatgaagactgcacgcgtgtgtgcacgcgggaagagacggtgc	5707
Db	1202	AGCATGCTTTCCCTACATCAAGACACGCTCCGAGTGTGCGACGTCGAGAGACATGTC	1261
QY	5708	tgaagcaggtgaggttgatcgtcgtgagaacatgaaagaagaagacacggagacgtgcgcttg	5767
Db	1262	TGACACAGGTGAAGTGGCCATTGAGACATGCGAAGAAAGAACCCGGAGCTGGCTTTTG	1321
QY	5768	ccaccgagcaggaacccacagatgtctaagaatgtctaaagattggtcttcaagggctctgag	5827
Db	1322	CCACGAGACAGAGACCTTCCAGATGCCAAGATGCTGCAGATGGTGTTCACAGSGTCTGTHG	1381
QY	5828	ggcccaacggttaacacaggttccctctgtaggttgcccaagtggttttaagcagaagatcccg	5887
Db	1382	GACCCACTGTGAACACAGGGTCCCTTGGAAATGGCCACAGGTTTTTGTGAGAGATCCAG	1441
QY	5888	aagaccccaagctcttcgcgcacatcaacaacaatttgcgctctgcttcaagagactcttgc	5947
Db	1442	AAGATCCCAAGCTTTCTCGAATCTACACAACAACTCCGGCTGTGCTTCAAGATTTCTGCA	1501
QY	5948	agaaatgtgagatgctctcgcggaaaaaataagccctgatltggccgcgcgaagaggt	6007
Db	1502	AAAATGTGAGAGATGCATGGAAGAAACAAAGGCTCTGATTGGCCACGACGAAGAGGT	1561
QY	6008	accacgctgagcttgagagcgaactacatcgcgcgcttgaggagagctctgcagcccctgcta	6067
Db	1562	ACCAACGCTGCTCCCACTCTCTGGACCAAGTTCCACACGCTCAAGAGACTCCATGACACA	1681
QY	6128	gagcagatlttcggaagcagagcctctcgagcccaacgaagaccaaagctgttacctagaga	6187
Db	1682	GATCAAGTTTTCAGGAAGGCTGACTCTTG --- ACAAGGCTTAAGAGCCACACCCGAAG -	1735
QY	6188	accaagaccgcgctcaagctgtctgtgctctcgaaggaagatctgcacctgttcccaatgc	6247



|||||  
Db 1736 ACCAGACGACACCCACCTGCTGCTTTCGAGATTAT----- 1778  
Qy 6248 gctgtgaggtgacacacactgtacttggggctggccctctgctccctgtgtcccatcgt 6307  
Db 1779 --AGTGCCTCCGCTGGAAGTCTAGAGATAGAGCCCTGCTGTCTGCACTGATGAT 1836  
Qy 6308 gtgcactgagctctccctcttcttaataatggtttataaagcaaaaaa 6367  
Db 1837 TCATAATTTGTTTTGTTGTTTAAATTAAAGAGCTTTATATACAAAAA 1896  
Qy 6368 aaaa 6371  
Db 1897 AAAA 1900  
RESULT 12  
AK055905 3327 bp mRNA linear PRI 31-OCT-2001  
LOCUS Homo sapiens CDNA FLJ31343 fis, clone MESAN100101, weakly similar  
DEFINITION to Rat trg gene product.  
ACCESSION AK055905  
VERSION AK055905.1 GI:16550749  
KEYWORDS oligo capping; fts (full insert sequence).  
SOURCE Homo sapiens normal mesangial cells (NHMC56046-2) CDNA to mRNA,  
clone lib:MESAN1 clone:MESAN100101.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,  
Fukuzumi,Y., Fujimori,Y., Komiyama,M., Sugiyama,T., Itle,R.,  
Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,  
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,  
Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,  
Matsushima,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,  
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,  
Nagahara,K., Masuno,Y., Nagai,K. and Isogai,T.  
TITLE NEDO human cdna sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3327)  
AUTHORS Isogai,T., Otsuki,T. and Sugiyama,T.  
JOURNAL Direct Submission  
SUBMITTED (24-OCT-2001) Takeo Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)  
NEDO human cdna sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan: CDNA full insert sequencing:  
Research Association for Biotechnology (RAB): CDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Eletechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: RAB and  
HRI.  
FEATURES  
Source location/Qualifiers  
1..3327  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MESAN100101"  
/cell\_type="normal mesangial cells (NHMC56046-2)"  
/clone\_lib="MESAN1"  
/note="cloning vector: pME18SFL3-primary culture, normal  
mesangial cells"  
54..2279  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB71042.1"  
/db\_xref="GI:16550750"  
/translation="MEFLSYPRQOHYLAIVTELAVIILDPABGLFGLHKVIMV  
HNLISHDSPRTSDPQIKAVMALVPLIGITMETVPLQIDTFETHNORFICAT  
IDYESGSMISQTVAMAIAGTSVPLTRGSFLITSTSGRHTTSAESSRLILCL  
LWYLNADFTVLQWFTDLSVLQINRLIDLIVCSGFEKKGKVERNNLSLFFKSK  
DMRAKLEALITLSIGAQEWVRRSRGQLERSPGSAFSGQENLRMRKDMTHMRNTEK

LKSRALIEHIALIDGNLATEANLIIDTLEIVQTVSTESKESITLGVAKVILASHM  
ACNOSAVYLOHCEPFRALVSKPELLFEFEQOCDLRLRHSISITRSHAS  
ASLYILMRONFEIGNFARVMQVMTSLISLVGTSONFNEEFLRKSITLTYAEDL  
ELRETFEPVOVDLFIENLHMLISDYEIMHEDPEMLIDLMYRIAKGTONSPDLRT  
WIONMAGKSHSERNNHAEAOCLVHSAALVAEYLSMEDKYLIPVCVFQONISVYLE  
ESAVSPDYSPDEEGICSGFPRESGLVILBDOAASFMAQVENVYVVLPIH  
EANRDKRKLSTIHKGLQEAESKTIYHVDGKRMSTYRNVCFYTKRGDLDDECFYKEP  
AIIKRLAEISRLEGFYGERGEDVENVIVKDSNPFVCKLDPNTHIFRLPMWSHTLTHM  
R"  
BASE COUNT 1046 a 640 c 711 g 930 t  
ORIGIN

Query Match 19.28; Score 1225.4; DB 9; Length 3327;  
Best Local Similarity 65.48; Pred. No. 7.7e-186;  
Matches 1873; Conservative 0; Mismatches 956; Indels 34; Gaps 4;

Qy 3245 cctcccaagctccacactctccagccagcccccagagctgacagcagctgcg 3304  
Db 1 CATCTCAGAGTCTGGATTCTTCTACGAATGTACAGACCAAAAGATTGCATAATGTTTG 60  
Qy 3305 aactgagtgacacatcccgagcagcactcctagctgggtctctgtaagagctgcg 3364  
Db 61 AATTATCGGTGCTTCCGCCAAGCATATTGTGGCAGACTGTGTAAACAGAGCTGG 120  
Qy 3365 cactggccctggaacctgaagctgaagggcattccctgttgcaagaagccatcagtg 3424  
Db 121 CTGTCAATTTAGACCTTATGCTGAAGGACTGTTGGATTCTCATAGAAATCTCATATA 180  
Qy 3425 ctgtgacagcctgtctatgtgccaatgacactgaccccccgatcgacgagggccacttga 3484  
Db 181 TGGTACACAAATTACTCTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTAA 240  
Qy 3485 agctctgtgtgacgagctgaagctgaagctgaagctgaagctgaagctgaagctgaag 3544  
Db 241 AGCTCGAGTGGCCATGTTGTATCTACCTCTGATGATGATGATGATGATGATGATGATGAT 300  
Qy 3545 gctctgacactctctgaagccagctgaagccagctgaagccagctgaagccagctgaag 3604  
Db 301 AGCTGATGATTTTACAGAAATCTCAATCAACGAGAAAGC-----CAATTTGTATAG 354  
Qy 3605 cagacacagaaagcgaaagggacatctgcggtacatcaaacccctctgtgacatgcca 3664  
Db 355 CCACGATGATATTATGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 414  
Qy 3665 ttgctgtgtgccccctagccccctgctcccgagcagatctcccaaggggacacaaagcg 3724  
Db 415 TCGCAGGGACATCGGCTCCCTCACTAACAGGCGTGGCAGATTCTCTCTCACTGTCACGA 474  
Qy 3725 ctcttcgagcagcgtgtgccccctctgctgaagtcagcagcctgtgtgagctgtgc 3784  
Db 475 GTGGCAGGCAACACACTACTCTTTACGCAAGATCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 534  
Qy 3785 tctgtgtgtgaacaaacacacgagccggtctctcgtgagcgtgtgagcagctgaacac 3844  
Db 535 TTTGGGTTCTCAAAATGAGATGAAGACAGTTCTACAGAAAGGTTTACAGATCTCTCAG 594  
Qy 3845 tccccagctggagagctcgtgtgagctgtgtgaacttgcctgagctgctgtgtgagta 3904  
Db 595 TCTTGGAGCTTAACCGCGCTATTAGATCTCTTATCTCTGTGTGTGTGTGTGTGTGTGTGT 654  
Qy 3905 aggggaaagagccttgaacgcatcaacagcctcaatcacaataatctctgataga 3964  
Db 655 AAGGGAAGAAAGTGTGTAACGAATGAATGATGATGATGATGATGATGATGATGATGATGATG 714  
Qy 3965 aggcgagctagaagaagcattctgtgtacatcagcagagctgcagaagaatggttcggc 4024  
Db 715 GAGCAAAAGCTTGAAGAGATATTCTTGGAGCATAGTGTCCAGGCAAGAAAGTGAACGGC 774  
Qy 4025 gaagctgtgagag-----agccggttgggaatccagacagc 4063  
Db 775 GAAGCCGAGACAGCTCGAGAGAGAGCCCATCTGGAAGTGTCTTGGAGTCAAGAAATTT 834

QY	4064	tgagctgaggaagacggtcaacacttgaaagaacacctcaagccgctgtagcaagacca	4123
Db	835	tgagctgaggaaggaatgatatgactcactgcggctgtaaaacacaaagaaaccttgacaatattaa	894
QY	4124	aggaatgaaatgaaacacgaagcccttggttgaaagggaaccctggcaaccgaaggaacactag	4183
Db	895	gagcagagatgtaaacacgaagacactgattgatgaaacactggcctacagaaacaaactttaa	954
QY	4184	tggtctcgaaacactgaaatcatcctgtgaaacggttgatgctttcaagaagcccgagaa	4243
Db	955	tcatttttagatnactttagaatttgctttcagaccctttctgttaacggatnccaanagaa	1014
QY	4244	ggctcttgagggcagtgctgaaagtgtgtgctgtacacccctgggcagtgcccaagtgccc	4303
Db	1015	gcattcttggtggagtgctttaaagtgctatctatcacagcattggcctgttaacaaatgcag	1074
QY	4304	tcctcttgcaacatgagccttgccaccccaagggcccttgctcaagttcccgagcgtgc	4363
Db	1075	ttttatctncaaacactgtttttgctncaacaaagaaaccttggtttcaaaagtgttcctgaactct	1134
QY	4364	tgctcgaaggagacaagagactgtgtgcgaacctgtgcctgaagtcctcctaagaactg	4423
Db	1135	tattttgaagaagacacagacacagacctgtgctgatttatgcctcagcgttctccgacactgta	1194
QY	4424	ggacgcgaatcaagacacatccgcacgaacgcagccagccctcgctgttaacccgtccatgcac	4483
Db	1195	gcagtagcatcggtatcaaatnaggtacaaagccagtcgctcctttacactnattgaagac	1254
QY	4484	agaactctcgagatctgcgcacaactttggcccggtgtgaaatgtcaagttcaacatgctctc	4543
Db	1255	aaaacttggaaattgggaataatnactttggccgggtttaaattgcaagtaaatgtcactat	1314
QY	4544	cgctccctgggtggggaacagcagaactcagttaaagaacacttgcaagcttaactcca	4603
Db	1315	ccctcttggtggcgaacatctcagaatttttatgaataattttaaagacgttcttaaga	1374
QY	4604	ccatccctacactatgctggaaggagacatggggctgtgggaacgcacccctcgaaagccag	4663
Db	1375	ctnattttgacatgatcgtgaagaaagatctggaaattggagaaacaaactttcctgattcagg	1434
QY	4664	tcacggagactgtatltcaacctgtcacatgtacctgtcagacacagttgaatltgaagaagac	4723
Db	1435	ttccagatctcgtgtttgcaattctccatnattgatgtctttctgatnctgtaaaatgaagac	1494
QY	4724	acagggagagaccctgaaatgtcatctgaacctgaatgaagaatltcccgggctcaccg	4783
Db	1495	accagagagatccttgaanaattgttgattgatnattgatnaccnaattgccanaagettaccgaa	1554
QY	4784	gtcaacggagacctgcgctgtacactggtgttcagaaatgtgcgggaagacacgagagactgtg	4843
Db	1555	ccctctccagatcttcgcactgacactcgtgttcgagaaatgtgcagaaacacgcaagcactcagaa	1614
QY	4844	gcaacacacgcgggaagcccgccacagtgcatgtgagtgagacgcgcgcgcctcgtagtgtagcc	4903
Db	1615	gcattatnctgcgaagctgcacactgtctagtccactcaccagacttggcttgaaattt	1674
QY	4904	tcgcctctgcgcgaagaccagcgccacactgcgcgttggtgcgtttcctctccagaacatct	4963
Db	1675	tgagatctgcctggaagaccgcggaataattcttctcttggaatgttaacattttcgaatattt	1734
QY	4964	catcaaaagtctcgaaggagtctgcgcactctccgaagaaatcctgtctgcgcgcgcgaaggg	5023
Db	1735	catctaatgttttgaagaatctgcggtctcagatnattgtgatttctcactgaatgaag	1794
QY	5024	gctctctgcctcgggaagcactltcaactgtagcttgggcttggtlaagttgtctgtaaacagag	5083
Db	1795	gtatctgccttggaataattacttttctgagtcagacagacttctgggaattactggaacaagcag	1854
QY	5084	ccggtacttcaacatggggcggtcctcaagagcggtagaatgtctcaagaacactcca	5143
Db	1855	ctgcttcttcttctatgctgctggcatgtatgaacacttattatgaattttacaatgactttaa	1914
QY	5144	tcaccatctctggaacccacacgtgactcaaaagacttggtgcgggtgtgcaagcgaacatgc	5203

Db	1915	TTCTCTATTCTGAAGCTAATCGGAGATGCAGAAAGAACTCTCCACATTCATGTTAAACTTTC	1974
Oy	5204	agagagcgcttcacaaagatcaltgacacagagtlcccgcttggagcgcglttgcggagcgt	5263
Db	1975	AAGAGAGATTCACCAAAATTTGTTCAATCAG-----GATGGTAACGGGATTTTGGCACCT	2028
Oy	5264	attccgcgcgtggccttcctcagcgcccaacttcggtgacgttgatgagcaggagtttgcgt	5323
Db	2029	ATTTCCTGTTGGTTTTTATGGACCAAGTTGGGGATTTTGGATGAGAACAGAAATTTGTTT	2088
Oy	5334	acaaagagccatcgcatacgcgaagcttgcagagatcctcacaccggtcvggaggttctaca	5383
Db	2089	ACAAAGGAGCCTCGAAATTAACCAACTTGGCAGAGATATCTCACAGATTCAGAGGAGATTTTACG	2148
Oy	5384	cggagagatttggcgcgcgcgtctgttgagatattatcaagaactctaccctgttgacaagt	5443
Db	2149	GAGAAATTTTGGAGAGAGATGTGGTTGAAGTATTAAGAACTCTAATCCGTGTGACAACT	2208
Oy	5444	ccaaagcttgcactcaaaaaggcctacatccagatcaatgltgtgtgaaacgctactttgata	5503
Db	2209	GTAATATTAGA-TCCTTAACAGCATATATTCATATTAATCCTATGTGACCCATCTTTTGACA	2267
Oy	5504	ccctcagcgcccaagagccgggtgtacactacttgcagccaaactatgvggctgcacatcc	5563
Db	2268	CATATGAGATGAGAGAGACAGATACACTTATTTGCACAAAATTAACAACTTCGTGCATTTCA	2327
Oy	5564	tgcttcgcagcgcttcaagccggatggggcggcacaggggagctgcgccgagcaacaaa	5623
Db	2328	TGTACTGTACACCCCTTTACTTATAGATGGCCGGCCCATGGGGAACTTCATGAACAAATTCA	2387
Oy	5624	agcgtaagacgcttgcacgaacccagcaagccctccctacatcaagaactgcatacgctg	5683
Db	2388	AAAGGAAGACCATTCCTGACACTGCTCATAGCCTTCTTATATTAACAAAGGGTCAAATG	2447
Oy	5684	tgctgcagccgggggagagacggtgcacgcgcagcttggaggtggcactcgagagacatgcaga	5743
Db	2448	TCACCTCAATAAAGAGAGATCATCTTATACCACTATTTGAAGTTGCTATTTGAGGACATGCACA	2507
Oy	5744	agaagacacggagcgcttgccttgcacccgagcaggaccacccagatgctaaagtctac	5803
Db	2508	AAAAGACACAGGAGTTGGCATTTGCAACACATCAGAGATCCCGAGACCCCAAAATCTTTC	2567
Oy	5804	agatggtgcttcaaggcctctgtatagggccacccgttbaacagaggtccctctggaggtgccc	5863
Db	2568	AGATGGTACTCCAGGAGATCTGTAGGCACACACAGTGAATCAGGGCCCTTTGCAAGTTGCC	2627
Oy	5864	aggtgtctttagcagagatcccggaagaccccaagctcttcgycatcacacaatgtgc	5923
Db	2628	AGGTTTTTTCGTGTGAATTAACCTAGTACGCCCAAAACCTTTATGACATCAATTAATTAATCTC	2687
Oy	5924	ggcctgtcttcaaggactctgcgaagaatgtgagagtgctgtcgcgaaaaataaagccc	5983
Db	2688	GACTCTGCTTTAAAGATTTTACTTAAAGAGTGTGAAGATGCTTAAAGAAAATTAAGAGCT	2747
Oy	5984	tgattggccgagacacgaagagatcacaccgttgagctvgagcgcaactactgcgcctgc	6043
Db	2748	TAAATTGGCCGGATCAAAAGAGATATCAAAAGGAACCTGAGAAAGAACTATCATCGCCTTA	2807
Oy	6044	ggagagctctgcagcccttgcacttaccagcgcgtcccccagct	6086
Db	2808	AAGAGGCCCTACAGCCACTGATCACAGAAAGATTCCTCACTT	2850
RESULT	13		
AX172978	AX172978	3253 bp	DNA linear
LOCUS	Sequence 105 from Patent WO0142297.		
DEFINITION	AX172978		
ACCESSION	AX172978		
VERSION	AX172978.1	GI:14597957	
KEYWORDS			
SOURCE	synthetic construct.		
ORGANISM	synthetic construct		

REFERENCE artificial sequence.  
AUTHORS Lu, P., Garman, J.D. and Candia, A.F.  
TITLE Clasp-3 transmembrane protein  
JOURNAL Patent: WO 0142297-A 105 14-JUN-2001;  
Arbor Vita Corporation (US)  
FEATURES  
source Location/Qualifiers  
1..3253  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="polynucleotide fragment"  
BASE COUNT 1009 a 633 c 705 g 906 t  
ORIGIN  
Query Match 18.9%; Score 1207; DB 6; Length 3253;  
Best Local Similarity 65.2%; Pred. No. 6,6e-183;  
Matches 1053; Conservative 0; Mismatches 955; Indels 33; Gaps 4;  
QY 3343 gggctctgctgacgagctgacgacgctgaccccgaaacctgagagctgaggggacatcctg 3402  
DB 4 GGACTGTGTTTTCAGAGCTGGCTGTATTTAGACCCCTGATGCTGAAGAGACTGTTTGA 63  
QY 3403 ttgcaaaaagggccatcagctgctgacagccgctctatgtgccaatgacacatgaccc 3462  
DB 64 TTGCATTAAGAAATCATATATGTCACAAATTTACTCTCAATCAGACACTCAACCCG 123  
QY 3463 ccgtacggcagagccacttgaagctcgtgtgcccagagcttacctgacactgcttcg 3522  
DB 124 CGGTACTCTGACCTCAGATTAAGGCTCGAGTGGCCATTTGTACTCTGATTTGAT 183  
QY 3523 attgcaaggatccttgcacagcgtcgaatgactgtctgtagggccaggtcagcgltca 3582  
DB 184 ATATATCATGGAAACTGTACTCAGCTGATGATTTTACGAACCTCAATCAACGAGGA 243  
QY 3583 agagctgacctgaatgcttgaacacagacagaggcagagggaacatltgcggtaaccatc 3642  
DB 244 AGAC-----CAATTTGTTAGCCACTGATGATTAAGAAAGTGAAGCGGAAGTATGATA 297  
QY 3643 aacccctctgtgtgcaatggccatctgtgtgtgccccctagccccctggtcccgggccagc 3702  
DB 298 AGCCAGACCGGTTGCATGGCAATCGCAGGACATCGGTCCTCACTAACAAGGCGCTGCG 357  
QY 3703 atctcccaagggtccacaaagcgtctgcgcagcgtgtgcacctctcgtcgaatcaagc 3762  
DB 358 AGTTTCTCTCTCACTGCAACGAAGTGGCAGCAACACATCACTTTACCAACAATCAAT 417  
QY 3763 cgaagctctgtgcgt 3822  
DB 418 CGAAGCTTTTGTATCTGTCTACTTTGGGTTCTCAAAAATGCAATGAATCAAGTTCTACAG 477  
QY 3823 cgtctggccactgaactgaacactccccagctggagcgtctgtgtgactgtgtacatc 3882  
DB 478 AAGTGGTTTACAAATCTCTCAGTCTTGACAGCTAAACCGGCTATTAATCTGGTTTATCTC 537  
QY 3883 tgcctgtgcgtcctttgaatcaaggaggaaaagagccttgaagcgtcaacacagcctcaaa 3942  
DB 538 TGTGTGCTTGTGCTTGAATTAAGGAAAAAGTGTGGAAGATGATTAATCTTGAC 597  
QY 3943 ttcaaaaactctgtgatatagaagcgcgctagaagacacatctgtgttaccatcgga 4002  
DB 598 TTTTAAGAAATCAAAAAGCATGAGAGCAAAAGCTTGAAGAAAGCTTATTTGTGGAGCAATAGGT 657  
QY 4003 gctcgaacaagaatgtgttcggaagtcgtg-----agaggagc 4041  
DB 658 GCCAGCGAAGAAATGATGCGGGAAGCGGACACAGCTCGAGAGAGCCCATTTGGAAGT 717  
QY 4042 ccgtttggagatccagagacgctgcgtgcggaagaagcgttcaacaactggaagcaaac 4101  
DB 718 GCCCTTGTGAAGTCAAGAAATTTGAGGTGAGGAAAGATATGACTCAGGCTCTCAAAAC 777  
QY 4102 tcagaccgctgtgacaagacaggaatgaaatgaaacacagagcctctgtgtggaaggagc 4161

DB 778 ACAGAGAAGCTTGCAAAATCAAGAGCAGAGATTGAACAGCAAGCACTGATTTGATGGAAC 837  
QY 4162 ctggcaaccgaaggcaagcctagtggtcttgacacactgtagatcatcgtgcagaagcgtg 4221  
DB 838 CTGGCTACAGAGCAAGCACTTAATCATATTTAGATATTAAGATTTGTTGACAGCCGT 897  
QY 4222 atgtttccagaagcccgaggagcgtcttggggcaggtgtgtagaagttgtgtgtacagc 4281  
DB 898 TCTGTAAACGGAAATCAAAAGAGAGCATTTCTGTGTGGAGTGTCTAAAGTGTCTACTACAGC 957  
QY 4282 ctggcagctggccagagtgccctctctcttgcaagcatgtgcctgcccaccccaaggagccct 4341  
DB 958 ATGGCTGTAAACCAATGTCAGTGTATCTACAAACACTGTTTGCTACAGAGAGCCCTTG 1017  
QY 4342 gtlcccaagtlcccgagcgtgtgtctgtagaaggacacggagcgtgtgtgcgcactgtgc 4401  
DB 1018 GTTCAAAAGTTTCTCGAACTTCTTATTTGAAGAGAGACAGCAAGTGTGTGATTTATGCG 1077  
QY 4402 ctgaggtcttcagacacgttggcagccgcatagaagccacatccgcaagcagcagcgcc 4461  
DB 1078 CTCAGGCTTCTCCGACACTGTAGCAGTACATGGTACAAATACGTTACACCCCAAGTGC 1137  
QY 4462 tcgctgtacactgctcatgcaagcagaactcgagatcgcgccacacacttgcgcgtgtgag 4521  
DB 1138 TCCCTTAACCTACTAATAGAGCAAAACTTTGAGATTTGGAAATACTTTGCCAGGGTTAA 1197  
QY 4522 atgaaagtcacatgtctctctgctccctgtgtggagagcagcagaactcagtgaaag 4581  
DB 1198 ATGCAGGTACCAATATGACTATCCCTCTGTGGGCACTCTCGAATTTTAAATGAAGAA 1257  
QY 4582 caccctgagcgttactccaacacatcctcaactcaactgtcgaaggagacatgaggcgctg 4641  
DB 1258 TTTTAAACGTTCTCTTAAGACTTAATTTGACATATGCTGAAGAAATCTGGAATTAGG 1317  
QY 4642 gacagaccttcgacagagcaggtccagagacttgatgtcaactgtgacatgacatgactgag 4701  
DB 1318 GAACAACATTTCCGTGATCAGTCCAGATCTGTTTCAATCTCAATATGATTTCTTCT 1377  
QY 4702 gaaacggttgaagaatgaagaaacacccagagagacccctgagatgtccatcgccatgtac 4761  
DB 1378 GATACTGTGAAGAAATGAAGAACACCGAGGATCTGAAGTTGATTTGATTTATTTAT 1437  
QY 4762 agaattgccggggtacacagagcttcaacagcactcgagctgtgactgtgtgcagaacatg 4821  
DB 1438 AGAATGGCAAGGTTTACCAACCTCTCCAGATCTGCGATTTGACTGTGTGCANAAATG 1497  
QY 4822 gccgggaagcagcggagctgtgggaacacacgcagggccgagctgcatgtgtgcacg 4881  
DB 1498 GCAGGCAAGCACTCGAAGCAAGCAATCATGCTGAAGCTGCACAGTGTCTTACCTCA 1557  
QY 4882 gccgcctcgtgtgtgagtaacctgcctcgtctgaagagcagcgccacactgcccgtggc 4941  
DB 1558 GCAGCACTTTGTGCAATTAATTTAGCATGCTGAGAGCCGGAATATCTTCCGTGGGA 1617  
QY 4942 tgcgttctctccaaacatcatccaaagctgtcgaaggaggtccgcacatctccgaagc 5001  
DB 1618 TGTGTAACATTTCAAGAAATTTTCAATCTTAATTTTGAAGAAATCTGCGGTCTCAGATGAT 1677  
QY 5002 atctgtgcgccagcagaggagcgtctgtcgcgggaagcactcactagactgtggggcgtg 5061  
DB 1678 GTGTATCTCCAGATGAGAGAGTATCTGTGGAATAATCTTTACTAGATCAGGACTT 1737  
QY 5062 gtaagggtgtgtggaacagcagcgcgctactcaacctcagcgcggtcctaaggcggtg 5121  
DB 1738 GTGGATTTACTGGAACAAGACGCTGCTTCTCTCTATGAGGTGCGCATGTAGAGCGACTT 1797  
QY 5122 aatgagcttacaagaacatcatcccatcctgtggaagccacacgtgtgactacaagaagcgtg 5181  
DB 1798 AATGAAGTTTAAAGTACTTATTTCTATTCATGACAGCTAATCGGAATGCAAGAAACTA 1857  
QY 5182 gccgcggtgtgacagcaactcgaaggagccttcaacaaatcatgacacagatgttcgcgc 5241  
DB 1858 TCCCAATTTCAATGATTAACCTTCAAGAACCAATTCAGCAAAATTTGTTCAATCAGATTCGCG 1917



```

Db      172 AGAAGACGCTTTAGATATGCGCTTCTTTGAGCTTCGGGAAACATGCGCCA 231
OY      2844 gcaactgtgcttggccagcgacatagacacaccccgaaactgtgccttcccgaagcct 2903
Db      232 GCACGTACATTAACATGACAAACGGGACAGTTTTCGGAGSACTGGTTTTCGACCGTTT 291
OY      2904 ccttgagacacatctgcttctgtgtgctctgtgtggtccttggaagtcacaccgtgt-- 2960
Db      292 CATGGATGACATTAATCTATTGTTAATGTGTCACCTCGGAAATTCGACGCTTTTAGT 351
OY      2961 -----ccacaagatgtgtagctggccgagcacctcaagcagcagccttgcttcttcc 3014
Db      352 AAAACCCAGAAAGAAATGAACAGCGGAAATGAAATGAACTCAGCCTGGCTTCTCTT 411
OY      3015 cagtgaaccttcttcccttgtgtggaacggcggtctgtcttccaccttggctccggacca 3074
Db      412 GTATGACCTTCTCCCTCATGATCGGGGCTTGTGTAACTCATCAGACATTTATG 471
OY      3075 caaggaagtgccagcgcgctccagtgctgctccctaalcacagcagcctgtgtaccctgc 3134
Db      472 CAGCGACGCTGACGCAAGCT-----CAGTAACCTTCCAAAGCTCATTTCCATGAG 522
OY      3135 catggaattcacccgcacatcgtgtgacccagcagcactacgtgaacctcaacctccctcg 3194
Db      523 GCTAGAGTTCTCTGAAATCCTCTGTAGCCATGACCATTTACTCAATCGAACT----- 576
OY      3195 ctgccccctgtcacctcagctgcctccctcctctgtctgtctccacacacctccagag 3254
Db      577 TTTTATTAATGAAATGCTGATGCTCTCCAAACATCTCTGTGCTTCCATCTTCCCGAA 636
OY      3255 ctccacctctccagcacaagccccgaaccccaaggtgacagcactgtgtgaaactgtg 3314
Db      637 CTCAAGCTCCTGTCACGACTTCCAGACACAGAAATGCGCAGCATGTTCATCTGACTTC 696
OY      3315 accatctccggaacgaacacttctcagctgtgtcctgtgtcgtcgtgacagctgtgcact 3374
Db      697 CGAGTACCGCGACGACGACTTCTCACCGGCTCTCTTCAAGAAATGCGCTGCTGCTT 756
OY      3375 cgaacctgaagctgaaggggacatccctgtgtgacagaagaagccaatgtgtgtgcag 3434
Db      757 GGATGCCGAAGGGAAGGAATCAGCAAAATGACAAAGAAAGAGCTGTCAATTCACAG 816
OY      3435 cctgtatgtgtgcacatgacatgacaccccgctacgcggaagcactgtgaagctgtcgt 3494
Db      817 CCTGCTAAGTCTTCACGACCTGACCCACGCTGTGTAACCAAGAGTGAAGGTCAAAAT 876
OY      3495 ggcgcagcttactctgcactgcttctcagatgtgcagygataccttgcacgctgtcat 3554
Db      877 CGCGGCCCTTACTACTCTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 936
OY      3555 cttgtcgtgaagcccaagctcagcggtcaagactgtgcctcaatgtctgactcagaacaga 3614
Db      937 CTTTACAGTTGCAATACTGCGAATACCGCA-----CCAGTGGCTCGGA 981
OY      3615 agcgcaagggaacatgtgcgggtacacacacccctctgtgacatgtgcacatgtgtgtg 3674
Db      982 TGAAGAACAGAAAGGAGCCGCTGCATTAAACAGAAATGTGCTGTGCCATACAGGAA 1041
OY      3675 ccccttagccccctggctcccgggcagcatctcccaaggggcacaacggcttctcgcg 3734
Db      1042 TAAATTTAAATTTGAAACAAAGGAATAGT-----GCTGTCTTCTTGGCCATTAAGCA 1095
OY      3735 aggcgtgacctctgtcgtacgaagccgagacctgtgcgtgtgtgtgtgtgtgtgtgt 3794
Db      1096 GTACAAACATGCTGAAACCGGACACTACTGCAACCTCATATGCTGCTCTCTGATCAT 1155
OY      3795 gaataaacacgagccggtcctctcagcgctgtggccaactgtacacttccccagct 3854
Db      1156 GAAAGAAAGCTGATCAGGCTCTCATTTAGGAAGTGTGCTGACCTCCATCAACGACGCT 1215
OY      3855 ggaagcgtgtgtgactgtgtacacttgcgcgtgtccttgaatataaaggggaaaaa 3914

Db      1216 CAACAGATTTAGATCTACTTTTCACTGTGTATGTTTGAATATAGGAAACAA 1275
OY      3915 ggccttgaacgcatcaaacgcctcaacatccaataatcttggatatagaagcgcggt 3974
Db      1276 GAGTTCTGACAAAGTCAATGACCAAGTCTTGCAGAACTCAAGGAGATCAAGGCGCGGT 1335
OY      3975 aagaagaagccattctgtgtacccatcgagactcgacagaagaatgtgttcgagatcgtga 4034
Db      1336 GGAAGAGGCTTGTCTCCGTGGGAGGGGACAGAGGGAGATGATCGCGCGGGCTCC 1395
OY      4035 gaaggaagcgtt-----ttggaaatccagaagaacgtgtgcgtgtggaagaagctcaaca 4088
Db      1396 AGGAACGACCGATTTCCAGGCGCTTAATTAATTAATTAATTAATTAATTAATTAAT 1455
OY      4089 ctggaagcaaacctcaagcgcgtgtgacaagaccagaatgaatgaatgaacagcagcct 4148
Db      1456 TTGGCGGCAAGCTAATGAGAGAGTAAATTAACAAAGCCGAGTTGATCAAGAAAGCTT 1515
OY      4149 gttggaaggaagcctgtgcaacccgaagcctagtgtgtctgtgacacatlgagatcat 4208
Db      1516 GATCACTGGCAATCTGCTACAGAAACACATTTAATCATCTGATATGCAAGAAACAT 1575
OY      4209 cgtgcagcagtgatgtcttccagaagcccggaagcgctcttggggcgagctgaagt 4268
Db      1576 TATCCAGGCGAGCTCGGCTTGACTGTAAAGACACCTGCTGGGAGGTCTTAAGGCT 1635
OY      4269 tgtcgtacagcctgtggcagtgccagatgtccctctctctgtcagcaltgcctggcac 4328
Db      1636 GCTGTGAAATTCCTGAACCTGTATGATGAGATACACCTACTGACTCACTGCTTTGCAAC 1695
OY      4329 ccaagagggcccttgttccaaagttcccgagctgtctgtctcggaggaggaacggagctgt 4388
Db      1696 ACTCCGTGCTCATGCTCCAAAGTTTGAAGCTTACTCTTTCGAAGAGGAGGTGAACAGTG 1755
OY      4389 tgcgcacccgtgtcgaagctcctacgacactgtgtgcacccgacatcaacatccgcac 4448
Db      1756 TTTGACACTATGCAACAGTCTGTGACACCTGACACGACACATGATGTACCCGGAG 1815
OY      4449 gcaagcagcgcctgtcgttactgtcatgtcagacagaacttcgaatcggccaaactt 4508
Db      1816 CCAGGCTGTGACACCTTACTCTCTCATGAGGTTTCAGATTTTGAAGCACACAGTAATTT 1875
OY      4509 tgcgcgtgtgaagatgacatgacatgtctctcgtccctgtgtgtggggaacgcagaa 4568
Db      1876 TGAAGAGATTAAGATCCAAATGATCAATGCTCTGCTGCTTGTGGGAAAGCACCGA 1935
OY      4569 cttaagtgaaagacactgtgcacgttcaactcaaaaacatctcaactatgtctgaagaga 4628
Db      1936 CTTTAAATGAAGACACCTGAGAAGATCTTGAAGCAATTTTGGCTTATTCAGAAAGGA 1995
OY      4629 catggtgtgtgtgacagacacttgcagagaggttccaagacctgtatgtccaacttga 4688
Db      1996 CAAAGCATGACAGATGACTCTTTTCCACCCGAGTGAAGAACTTCTCTTAATCTGAA 2055
OY      4689 catgatccctgaagcacaacggttgaagatgaaggaacacaggaagcctctgagatgat 4748
Db      2056 TAGCATCTTAATATGACACAGTGAATGAAGGAATTTAGAGAAAGTCTTGAGATGCTTAT 2115
OY      4749 cgaactcatgtacagaattgtccgggtctaccaaggttcacacggacttgcgttgcactgt 4808
Db      2116 GGATCTCATGTACAGAAATTTGCCAAGATTACAGGACATCTCCTGATGCGGCTGACTG 2175
OY      4809 gttgcaaacatgtgcgcgggaagcagcgaggtgtggaacacagcgcggagcgcccgagt 4868
Db      2176 GCTTCAGAAATGAGGAGAAACACACAAAGAAAGATGCTTACAGGAGGCTGCATGTG 2235
OY      4869 catgtgtcacggtgcgcgcctcgtgtgtgagttacactgcctctcgtgcagacacagcca 4928
Db      2236 CCGTGTGACGCGCGCTGGTTAGTGGCTGAGATATCTGAGCATGCTGAGAGACACACTA 2295
OY      4929 cctgcgcgtgtgtgtgtgttctcctcagaacatctcaacagtcgtatagaaggtcgc 4988
Db      2296 CCGTGCCTGTGGCAGTGTACAGCTTCCAGAAATATTTCTTCAATGTGCTGAGAGAGTGT 2355
```

```
Qy 4989 catctccagacacatcctctgcgcgacgagagaggtctctccggaagcaacttcac 5048
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2356 GGTCTCTGAGAGACACCTGTCACTGACGAGATGGGGTGTGGCGAGGCCACTTTCAC 2415
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5049 tgaagctggagctgtgaggtgtcttgaaacagcagcgtacttaacatggcgagct 5108
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2416 CGAGAGTGGCCTGTGTAGGCTCTCTGAGACAGCGCGGAGCTCTTACAGCAGGGAGCTT 2475
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5109 ctacagagcggtgaatgaaggtctacaagaacctcaccatctctgaagcccaagctga 5168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2476 ATATGAGACAGTTAATGAGGTATACAAAGCTGGTCACTCCCATCTGAAAGCGCATGAGA 2535
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5169 ctacaagaagctggtccggtgtgacagcgcaactgcacgagagcgcttcaccaagatctga 5228
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2536 ATTCCGAGACCTGACACTCACTCAGCAGCAAGCGCAGAGAGCCTTGGACAGCATCGTTAA 2595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5229 ccagaggtccggtgtgagagcggtgtctggagacgtatctccggtgtgagctctcaagcgcg 5288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2596 CAGG-----GATCATNAGAGAAATGTTTGGAACTTATCCGAGTTGGTTCTTTGGATC 2649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5289 ccacttcgtacgtgagatgaagcagaggttgtgtacaagaagcatalcatcagaagct 5348
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2650 CAATTTGGGGGATTTGGATATACAGGAATTTGTCTCAAAAGAGCCGCAATTTACAAAGCT 2709
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5349 ggcagagatctcaacccggtgtgagaggtctcaacgagagagatttgcgcgagcgctgt 5408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2710 TCCGAGATCTCACTAGACTAGAGGCAATTTATGTCATATGTTTGGTGCAAGAAATTTGT 2769
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5409 tggagattataaagaactcttaacctgtgagacatgcaagctgactcaaaagcgcta 5468
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2770 GGAAGTATTAAGAACTCCCTCTGTGACAAACCAAGTTGGATCTTAACAGGCTTA 2829
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5469 calccagatcaactgtatgtgaacccgtacttga taactaagactcaaaagcagcggtgag 5528
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2830 CATACAGATCACTTTGTGAGAGCCCTATTGATGAGTATGAGATGAAGACAGGCTCAC 2889
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5529 ctactttgacgcgcaactatgaggtcttgacacatctccggttctgacgagcgctcaagcgga 5588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2890 ATACTTTGAGAAAGAAATTTCAACCTCCGAGTTTATGTATGATACACACCCCGTTCACCTGGA 2949
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5589 tgggcgcgcacacgggaggtctccgcgagcaacacagcgtgaagcgctgtcagccagca 5648
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2950 GGGGCGGCGCTCGGGAGAGAGCTGCATGACAGTACAGAGAAACAGACTCTGACCACTAT 3009
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5649 ccacgctccctccatcaacaagctcgatccgltgtgtccaccggagagagacggtgct 5708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3010 GCACGCTTCCTCCATCAATCAAGACACAGATCACGCTCATCAGAGAGAGAGTGTGTTT 3069
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5709 gaagccaggtgaggtgtgcctatcgagacatgcatgaagaagaacacgagagctgtgcctttgc 5768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3070 GACACCGATTGGAATTCCTCCATTGAAGACATGAAGAAAGAACCCCTGCAGTTTACAGTTGC 3129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5769 caccgagcaggaaccacacagatgaatgtctacagatgtgtctcaagcgctctgag 5828
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3130 CATTAACACAGAGAGCGGCTGTATGCAAAAGATGCTTCAGATGTGTCTGCAAGAGCTGTGTGG 3189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5829 gccacacgtgaacacaggtgtccctgtgaggtgtgcccaggtgtgttttaagcagagatacccgga 5888
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3190 AGCTACGTATAATCAGGAGACACTGGAAGTACGCCAAGCTTTTGTGGCTGAATATCTCTGC 3249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5889 agaccccaagctctctccggaatcaacaacaatgtgcgtctgtctcaaggaactctgcaa 5948
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3250 TGAATCCAAAACCTTATGTGACATCAACAAAGTGTGAGGTTATGCTTTAAAGAAATTCATCAT 3309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 5949 gaattgtgagagatgcgtctgagaaanaataagcgctgtatgtggccgagcagagaagagta 6008
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3310 GAGATGTGTGTAAGCTGTAGAGAAAACAAAGCGTCTCATATACGGCAGCAGACAGAGGAATA 3369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 6009 ccacccgtgagctgtgagcgcacactactgtcgccctgtcgagagagctctgcagccctgtctac 6068
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3370 TCACAGGAGAACTCAAAAAGAACTATTAACAAGCTTAAAGAGAGAACCTCAGGCCAATGATCGA 3429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Qy 6069 ccagcgctgccccagagct 6087
Db 3430 GCGGAAAATTCAGAGACTG 3448

RESULT 15
AX174571
LOCUS AX174571 4026 bp mRNA linear PAT 04-JUL-2001
DEFINITION Sequence 3 from Patent WO0142296.
ACCESSION AX174571
VERSION AX174571.1 GI:14625747
KEYWORDS
SOURCE human.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 4026)
AUTHORS Lu,P., Garman,J.D. and Candia,A.F.
TITLE Clasp-5 transmembrane protein
JOURNAL Patent: WO 0142296-A 3 14-JUN-2001;
Ardor Vita Corporation (US)
FEATURES
source
1..4026
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="preliminary human CLASP-5 cDNA sequence"
<1..2970
/codon_start=1
/protein_id="CAC43926.1"
/db_xref="GI:14625748"
/translation="SPTMADTAPSPSPSPSSONSSSCSSSFODOKIASMPDRSRVA
STSSPGLITPELAALDABEGISEVQKRAVARIHLSHSDHDPKCYKPEYKXIA
ALYLPVGIILDALPOLCDFTVADTRNYRISGSDDEBGALITONVALAIAGNENL
KTSGLIVSLSPYKOYNMLNADDTYTNQVLOKSDYVARAEALIREGARGEMRRRAGN
DLRFIVLCEFEYKQKSDKSTQVLOKSDYVARAEALIREGARGEMRRRAGN
DRPGLDNENLRMKKEORTHMROANEDKTKTAELDOEALISGNLATFAHLIIDMOENI
IOASALNDCKDILGVLRLVNSLNDOSTYLTTHGFATRLALIAKFGDLDFEEVE
OCPLDQHOVLHHCSSMDVYRSQACATLYLIMPRSGATSNFARVMQVMTSLSYG
RADPFNEHLRSLRTILATSEEDTQAMOTPFPOVEBELCNLSILYTVKREFOE
RPEMLMDLMRIAKSYQASPLRLTLWQANAEKHTKCTTEAMCIVHAALVAETL
SMELDSYLPVGSYSPONISNVLSESVSEBDLSPDEBDCAGQVFTESGLVGLLEQ
AALFSTGLGLEVEYKLVIPLELHAREFRKLTLSHSLQAFQSIYVKDKRMFG
TVFRVGFGEKFGEDLDEDEYKPEPATTKTPEISHRELAQYGCFAEVEYLDKSTP
VDKTKLDPNKAYIOITPEYPPFDEYKEDVITYEKENLNRRRMYTTPPLLEGRPGE
LHEDYRRNYVLTTHAFPIYKTRISVLOKEFVUTPEVIAIDMKKTILOLAATINDE
PPDAKMLQMYLQSGVATVNOGPLEVAOVFLAEIPADPKLIRHNKRLDCKEFLMK
GEAVEKMKRLITADOREYQDELKKNYKLENLPMIERKIPELYKPIFEVESQKRDS
FHRSSFRKCTOLSGS"
BASE COUNT 1161 a 898 c 957 g 1010 t
ORIGIN
```

Query Match 16.2%; Score 1031.8; DB 6; Length 4026;  
Best Local Similarity 61.5%; Pred. No. 5.6e-155;  
Matches 1771; Conservative 0; Mismatches 1072; Indels 36; Gaps 6;

```
Qy 3218 cgcctccctctgtgtctctccaccactccacagatccacacttctcagcccaagccc 3277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 23 CTCCAACATCTCTGTGCTCTTCATATCTTCCAGAACTCAAGTCCGTGCTCGAGCTTCC 82
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3278 cggagcccaagtgacagagcagcagtgctcgaactgagagccact--ccgagcagagcact 3335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83 AGGACCAAGAAAGATCCCGACAGATGTCGAGACTTCCAGAGTACCGCCGACGACACTT 142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3336 cct-agcttgagctctctgtgacgagctgtgacactgtgcaactgtgaagctgaaggg 3394
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 CCTACCGGGCGCTCTCTTCAACAGAACTGGCTGCGCTCGAATGCCAAGGGAAGGAA 202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3395 catctctgttgacaagaagccatcagtgctgtgacacagcctgtatgtgtccatgaca 3454
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 TCAGCGAAGTACAAAGAGAGCTGTCAAGTATTCACAGGCTGTGATGTTCTCACGACC 262
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

[illegible]

Dp	1332	TACACATGTCCCTGGCATCTTTGGTGGGAAGCACACACTTTTAATGAAAGCACTTGA	1381
Qy	4589	gaagctaaccaaaacacatccctaacctatgctgagagagaaatggygctgagacaga	4648
Dp	1382	GAACATCCTTGAGAGACAAATTTTGGCTATTTCGAAGAGAGACACACCATGACATGACTC	1441
Qy	4649	ccttcgagagagaggtcccgaggaactgattgtaacctgcacatgatcctctgacgacg	4708
Dp	1442	CTTTTCCACCCAGGTGGAGAACTTCTGTAAATTCGAAATGACATCTTAATGACACAG	1501
Qy	4709	tgaagatgaagaaacacacaaagagagacctgagatgctacatcgactcatgtaacaaatg	4768
Dp	1502	TGAAATATGAGGGAATTTACAGGAAGATCTGAGATGCTTAATGATCTCATATGACGAATTG	1561
Qy	4769	ccgaggtctacaggggtctcaacggaaccttcggtctgactggtgttcagaacaatggcggga	4828
Dp	1562	CGAAGATTTACACAGGACTTCCTCGATCTCGGCTGACCTGGCTCCAGAACATGGCAGAGA	1621
Qy	4829	agacacgagagcttgggaaacacacgacggagccggccagtgatggtgtaacgcgccgcc	4888
Dp	1622	AACACACCAAGAGAAGTGTCTACACGAGAGCTGCCATGTGCTGTGTCACGCCCTGTGCT	1681
Qy	4889	tctggtctgtagtacctcgccctgctcgagagacagcgccacctggccggtggtctgctt	4948
Dp	1682	TAGTGGCTGATATCTGAGATACCTGTGAGAGACACACAGCTACTCTCGGTGGCGAGTGTCA	1741
Qy	4949	ccttcagaaacatctcatccaaagctgctgagagagtcgcacatctccgaagaaacctct	5008
Dp	1742	GCTTCCAGAAATATTTCTTCCAAATGTCTGTGAGAGATCTGTGTCTCTGAGAGACACCTGT	1801
Qy	5009	cgcccgacgagagggctcttcgtccgggaaagcaacttcactgaacttgggcttggtaagt	5068
Dp	1802	CACCTGACGAGAGATGGGGGTGTGGCGAGGCCACTACTTCCAGGAAGTGGCTGTGTA GCC	1861
Qy	5069	tgcctgagaaagcgagccggtacttaccatgaggcggtctlaagagcggtgaatgaag	5128
Dp	1862	TCTGTGGAGCGCGCGGGAGACTTTTACGACGGGAGGCTTATATGAGACAGTTAAATGAG	1921
Qy	5129	tctcaagaagacctccatcccatctcgtbgaagccacacgctgaactaaagaactgcccgg	5188
Dp	1922	TCTACAAAGCTGTGATCCCATCTTAAGAGGCCATTCGGAATTCGGAGACTGACACTCA	1981
Qy	5189	tgcacggcaactgcgcaggaagccttcaaccaagatcatgaccagaagttccggtctggagc	5248
Dp	1982	CTCACACCAAGCTGCAGAGAGCCTTGACAGATCGTTTAAACAG-----GATCATTAAGA	2035
Qy	5249	gcggtctcggagagatcttcgcgctggggtcttcaagtcggccaccttggagactgagt	5308
Dp	2036	GAATGTTTGGAACTTACTTCCAGTTGGTTGTTTGGATTCCAAAATTTGGGGATTTGGATG	2095
Qy	5309	agcagagatttgttcaaaagagacatcgatcaogaagcttgcagagatctcaacccggc	5368
Dp	2096	AACGGAGTTTGTCTCAAAAGAGAGCTGCATATACCACAACTCTCGAATCTCATATGAC	2155
Qy	5369	tggagagagttctcaacgagagatcttggcgagcagctgttgtagatatataaagactct	5428
Dp	2156	TAGAGGCAATTTTATGTCATGTTTTGGTGAGAAATTTTGGAAGTATTTAAAGACTCCA	2215
Qy	5429	accctctggaacagtcacagacttgaactcaaaaagcgcttaccatccagatcaagtatg	5488
Dp	2216	CTCTGTGGACAAACCAAGTTGGATTCCTAAACAAAGCCTCATATGACATCATCTTTGTGG	2275
Qy	5489	aacgcgtacttgaactcagctcagctcaagaacgggtgaacctactttgaccgaactatg	5548
Dp	2276	AGGCGCTAATTTGATGTAGATGTGAAGACAGAGGCTCACATCTTTGACAAATAATTTCA	2335
Qy	5549	gagctcgacattccctgctctgacacgcgttcaacgagatggcgagcgacacaggggagc	5608
Dp	2336	ACCTCCGAGAGTTCAATGTAACACACCCGTTTACCCCTTGAAGGGCGGCTCTGGGGAGAGC	2395
Qy	5609	tgcgcgagaaacaagcgtaagacgtgtctcgaacgacacgacgcttcccttaatat	5668

Db	2396	TGCATGAGCAGTACAGAAAGAAACACAGCTCTACCATATGACGCCCTCTCCCTATCATCA	2445
Qy	5669	agactcgcatccgtgtgtgccaccggyaggaacgylgtcgaagccagtgaaglybgca	5728
Db	2456	AGACACAGATACAGCGATCATCCACAAAGAGAGACTTTGTTTGACACCGATTGAATTGCCA	2515
Qy	5729	tcgagagcatcgaggaagaagacacggyagcttgccttttggcacccgagaggaaccacag	5788
Db	2516	TTGAGACATGAAGAAAGAAACCCCTCACTATGACAGTTGCCATTAAACGAGAGCGCGCTG	2575
Qy	5789	atgcataagatgtacagatgtgtgttcacagagctctgtlaagggccacacg tgaacagagtc	5848
Db	2576	ATCGAAAGATGCTTCAGATGTGTCTCGAAGGCTGTGTGGAGACTACTTAATCAGGAC	2635
Qy	5849	ccctgtgaggtgtgcccagagtggttttagcaagagatccccgaagaaccacgtctccgac	5908
Db	2636	CACGTGAGGTAGGCCCAAGAGCTTTTGGCTGAATTCCTGCGATCCAAACCTCATGAC	2695
Qy	5909	atccacacaattgcgagctctgtcttaagaagattctgcagaagaatgtagatgcgtgc	5968
Db	2696	ATCCAAACAAGTTGAGGTTATGCTTTTAAAGAAATTCATCTGTGAGATGTGTGAACCTGAC	2755
Qy	5969	ggaanaaa laaagccctgattgtggcgcggaacagaagagatlacacccgttagcttgaagcga	6028
Db	2756	AGAAAAACAAGCGCTCATCATCAGCGCAGACACAGGGGAATATCAGCAGGAACCTCAAAAAGA	2815
Qy	6029	actactgcgccttcggyagagctctgcagagccctcttaaccagcgcttgcacagctg	6087
Db	2816	ACTTAAACAAGCTTAAAGAGAACCTTCAGGCCATATGTCAGGGGAAATTCACAAACCTG	2874

Search completed: October 5, 2002, 07:35:44  
Job time: 18981 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 00:56:57 ; Search time 6621.52 Seconds  
(without alignments)  
12988.347 Million cell updates/sec

Title: US-09-736-968a-1

Perfect score: 6372  
Sequence: 1 gacgcgagagacatgctgc.....aaagcaaaaaaaaaaaaaa 6372

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: \*  
1: em\_estdb: \*  
2: em\_esthum: \*  
3: em\_estlin: \*  
4: em\_estmu: \*  
5: em\_estov: \*  
6: em\_estpl: \*  
7: em\_estro: \*  
8: em\_hic: \*  
9: gb\_est1: \*  
10: gb\_est2: \*  
11: gb\_hic: \*  
12: gb\_gss: \*  
13: en\_gss\_hum: \*  
14: en\_gss\_inv: \*  
15: en\_gss\_pln: \*  
16: en\_gss\_vrl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	1968.2	30.9	2710	11	AK016495 Mus muscu
2	1486	23.3	1510	11	BC020473 Homo sapi
3	865.4	13.6	1507	11	AK010755 Mus muscu
4	797	12.5	1112	10	BM473590 AGENCOURT
5	768.4	12.1	2694	11	AK004816 Mus muscu
6	762.4	12.0	796	10	B1770146 Mus muscu
7	758.4	11.9	815	10	B1831979 Mus muscu
8	751.4	11.8	927	10	B1830302 Mus muscu
9	738	11.6	770	10	B1830310 Mus muscu
10	735	11.5	794	9	A1198543 q49f11.x
11	725.2	11.4	793	9	AM027453 w59603.x
12	720.4	11.3	709	10	BE569035 602184383
13	683.4	10.7	745	10	B1822534 602624767
14	677.4	10.6	745	10	B1822534 602624767
15	673.8	10.3	662	9	A1653716 w536e07.x
16	659.4	10.3	654	10	BE857715 7946903.x
17	644.4	10.1	654	10	BE857715 7946903.x

C 18	644.4	10.1	655	9	A1769655
C 19	630.8	9.9	644	9	A1582283
C 20	608.2	9.6	640	10	BF056301
C 21	608.2	9.5	769	10	B181383
C 22	607.8	9.5	612	9	AM516592
C 23	607.6	9.5	1311	11	BC018599
C 24	602.2	9.5	630	9	A1800536
C 25	599	9.4	600	10	BM023316
C 26	597.6	9.4	702	10	BM023316
C 27	597.4	9.4	685	10	BF062076
C 28	582.6	9.1	777	10	B1078974
C 29	581.6	9.1	612	9	AM027279
C 30	580	9.1	588	10	BM023626
C 31	576.6	9.0	584	10	BM126124
C 32	576.6	9.0	632	9	A1765069
C 33	572.4	9.0	938	10	BF980727
C 34	570	8.9	778	10	B159195
C 35	569.4	8.9	589	9	AM168286
C 36	569	8.9	569	9	AM771566
C 37	568.6	8.9	842	10	B1854143
C 38	568	8.9	576	9	A1950836
C 39	567.4	8.9	569	9	A1694380
C 40	565	8.9	577	10	BM126416
C 41	564.8	8.9	569	10	BE301939
C 42	558.8	8.8	570	10	BF115265
C 43	547	8.6	547	10	BE568737
C 44	545.6	8.6	555	9	A1913163
C 45	544.8	8.5	718	10	BE311583

## ALIGNMENTS

### RESULT 1

#### AK016495

#### LOCUS

#### DEFINITION

AK016495 2710 bp mRNA linear HTC 19-JAN-2002  
Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4931431C02:homolog to KIA1395 PROTEIN (PRAGMENT), full insert sequence.

#### ACCESSION

#### VERSION

#### KEYWORDS

#### SOURCE

#### ORGANISM

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### MEDLINE

#### PUBMED

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### MEDLINE

#### PUBMED

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### MEDLINE

#### PUBMED

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### MEDLINE

#### PUBMED

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### MEDLINE

#### PUBMED

#### REFERENCE

#### AUTHORS

#### TITLE

#### JOURNAL

#### MEDLINE

#### PUBMED

#### REFERENCE

#### AUTHORS

AK016495 2710 bp mRNA linear HTC 19-JAN-2002  
Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4931431C02:homolog to KIA1395 PROTEIN (PRAGMENT), full insert sequence.  
AK016495.1 GI:12855259  
HTC; CAP trapper.  
Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library  
clone:4931431C02.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
10349636  
2 (sites)  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
2049374  
11042159  
3 (sites)  
Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishikawa, K., Katsunuma, T., Tashiro, H., Itoh, M., Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishwa, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Makiyama, K., Tanaka, T., Matsura, S., Kawai, O., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
Sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)



Db	840	CCTCCCTCTACCTGCTTAATGCGCCAGAACTTGAGATTGGCCATTACTTTCGCCGTGTGA	899
QY	4520	agatgtgagtcacacatgtctctctgtccctgggtgggagcagcagaacttcagtgaa	4579
Db	900	AGATGCTGTGACCAATGTCTCTGCTGCTCCCTTGTGGGGCAAACTCAGAACTTTAGTGAAG	959
QY	4580	agagccctgagcgtttactcaaaacactactactatgtctgtagagagacatgggctgc	4639
Db	960	AGCAATTTGAAGAAAGTCCCTTAAGACCATCTTGAACCTACCGAAGGAGGACATAGGGCTGA	1019
QY	4640	ggagcagacaccttcgcagagcaggttcacagacctatgtctcaactctgacatga ttcctga	4699
Db	1020	GGGACAGCACACTTCGGTGAGCAGGTCCAGACOTCATGTTCAACTCGACATGATCTCGA	1079
QY	4700	cggagacagtgtagatgtgaaggaacaccagagagagccctgagatgtctatgaactcatgt	4759
Db	1080	CAGACACGGTGAAGATCAAGTAAGAAACACAGAGAGACCTGATGCTCATGGAACCTCAATG	1139
QY	4760	acagaatttcgcccgggtctacacaggtctcaacgaccttcgctctgtagacctgttgtagaaca	4819
Db	1140	ACGAATATTGCTCGGGGATATACCAAGGCTCTTCAGAACTGGCGGTGACATGGCTCGAGAAAC	1199
QY	4820	tggcgcgggaagcagcgcgagctgtggtgcaacacagccgaggtcgccagtgtcatgtgtcaag	4879
Db	1200	TGGTGTGGCAAAATGACGACCTGGGCAATATGACAAAGGCTCTCCACAGTATGGTCAACG	1259
QY	4880	cggcgcgccctcgtgtgctgtgagttacctcgccctcgtctgtaggaccagcgcaacctgcctgtg	4939
Db	1260	CCGCTGCCCTGTGGCCGATTAACCTTGCCCTCTCGAGGACAGTGGCCACCTGCCCTGTGG	1319
QY	4940	gctgcgtttcccttcagaaacatctcatcaacgctgtcagtagagatcgccatctccagag	4999
Db	1320	GCTGTGTGCTCTTCACAAATGCTCTCATCCAACTGCTGGAGAGATCTGCATCTTGATG	1379
QY	5000	acatcctgtgcgcccagcagagagaggtcttcgtctccggagagacttcactagctgtgagggc	5059
Db	1380	ACATCTCTGTGCCAGATGAGGAGGGGCTTCTGCTTAGGAAAGAACTTCACAGAACTGGGCTC	1439
QY	5060	tgtagtaggtgtgcgtggaacagcagcgcgctacttcaacatgggcgggtctacagagcg	5119
Db	1440	TGGTGGGGCTCTCTGGAGCAGGCGGCGGCTACTTCCATCGATGGGTGTGTATGAAGCGG	1499
QY	5120	tgaattgaggtcttaagaacctatcccatcctctgtaaacccacccgttactataagaagc	5179
Db	1500	TGAACGAAGTTTAAAAAACCTTATCCCATCTTAAACCCACAGACATCAAGAAAGC	1559
QY	5180	tgagcgcggtgcagcgcaaacctgcagagagccttccacaaagtcatgcacagagttccg	5239
Db	1560	TGGCGTGGGTGTCACGGGAAATCGAAGAGGCTTCAACCAAAATTTATGCACAGACCTGTG	1619
QY	5240	gctcgggagcggtgtctcgggagcgtatctcccggtgggtcttcaagcgccgcaactcgggt	5299
Db	1620	GCTGGAGACCGGTATTTTGGGACATATTTCCGATGGGCTTCTATGAGCACATTTGGTG	1679
QY	5300	acctgtagtagacagagattgtgttcaagaagagcatcgatccagagcctgtggaagagact	5359
Db	1680	ACCTGTGATGAACAGAAGTTGTGTATCAGGAACCGTCAATCAAGAGCTTGCAGAGATCT	1739
QY	5360	cacacgcggtcgtgagagattcttcaacgtagagattgtgcagacagctcgtttgatatatca	5419
Db	1740	CACACCGGCTGGAGAGATTCTTATACGAAAGGTTCCGGGATGATGTGATGAGATCATCA	1799
QY	5420	aagactcttaccctgtgtgacaagtccaagctgttactcaaaaagcctacatccagatca	5479
Db	1800	AAGATTCTTAACCCGATGAGACAAAGTCAAGCTGGAGCCACAGAAGGGGTATCAATACGATTA	1859
QY	5480	cgtagtgtgaacccgtacttggatacccaagagtcacaaagcagccgggtgtacctcttgacc	5539
Db	1860	CCATGTGTGAGACCACATTTGTGAGACCTTATAGCTCAAGGATGGGGTGAAGCTACTTCGATC	1919
QY	5540	gcaactatgggtcttcgacatctccgttcttgcagcgcgttcaacgcgcgagtggtggcgac	5599

Db	1920	GSAACTATGAGGCTGGCGGGCCCTTCCTTCCTGCAACACCCTTCAACACGATGAGCTGGGC	1979
Qy	5600	acggggaagctgcgcgagcaacaacagcgtaagcgtctgaacacgcgacccagccttc	5639
Db	1980	ACGGAGAGTGTGGCCGACAGCACAAMCCCAAGACGCTCTCAGACACGAGCATCCCTTC	2039
Qy	5660	cctaataaagactcgactccgtgtgtgtccacgcggaggaagcgtgtgtgaacgcagtgtg	5719
Db	2040	CCTACATCAAGACACGCATCCGAGTGTCCACCGTGAAGSAAACATGCTGTGACACAGTGG	2099
Qy	5720	agggtgcactcgaggagacatgcagaagaagacacggagacttgcctcttgcacccagcagg	5779
Db	2100	AGGTGGCATTGAGACATGCAGAAGAGACCCGGAGACTGACCTTTCACCCAGCAGG	2159
Qy	5780	accacccagatgctaaagtctacagatggtgtgtctcaagctctctgttaggcaccgcgtga	5839
Db	2160	ACCCCTCCAGATGCCAAGATGCTGCAGATGTTCTCCAGGTTCTGTGGACCCACTGTGA	2219
Qy	5840	accagagtcctccctgtagagtgtgcccaggtgttlltagcagagatcccggaagaccccaagc	5899
Db	2220	ACCAAGGTCCTTGGAAAGTGGCCAGGTGTTTTGTGTGAGATGCCAGATGCCAAGTCCCAAGC	2279
Qy	5900	tcttcgcgcatcaacaacaattgcygctctctgtcttaaggaactctgcgaagaatgttaggg	5959
Db	2280	TCCTTCGGAATACACAACAACACTCGGCTCTGTGTTAAAGATTCTGCAAAAAGTCCGAGG	2339
Qy	5960	atcgctcgggaaaaataaagagcccgatlttggtgcggaacagaagagatcaacacgctgac	6019
Db	2340	ATGCACATGGAAGAAAGAACAGGCCCTGATTTGGGCCACAGACAGAAGSAGTACCAACCGGGAGC	2399
Qy	6020	tggagcgcaactactgcgcgctcggtgggaggtctgtcacgcccctgtactccaagcgctgc	6079
Db	2400	TGGAGAGTCTACTATATAGCCGCTCGGGGAGGCTCTCCACGCTCTGTGTCCCAACCTGTGC	2459
Qy	6080	cccgagctgactgacacccaccccccacccgcgcctcaagaaactcccttgaacagagcaagtcc	6139
Db	2460	CCCAAGCTGCTGGACCAAGTTCCACCAAGCTCTAGAGAGCTCCATGAACAGATCAAGATTCA	2519
Qy	6140	gaagagcagacactctgagaccacaagaagcaagctgtactatagagaagacaagaccgg	6199
Db	2520	GGAAAGGCTACCTCTTG-----ACAAGGCTAAAGAGCCACACACAGAAAGACAGCACCCAC	2574
Qy	6200	gctcaagctgtctgtgc 6216	
Db	2575	ACCTACGCTGCTGCTGTC 2591	

RESULT	2			
LOCUS	BC020473/c			
DEFINITION	BC020473	1510 bp	mRNA	linear
ACCESSION	BC020473			HTC 03-JAN-2002
VERSION	BC020473.1			
KEYWORDS	HTC.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
TITLE	1 (bases 1 to 1510)			
AUTHORS	Strausberg, R.			
JOURNAL	Direct Submission Submitted (03-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>			
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgapbs-r@mail.nih.gov">cgapbs-r@mail.nih.gov</a> Tissue Procurement: ATCC CDNA library Preparation: Life Technologies, Inc. CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center			



AUTHORS	Carninci,P. and Hayashizaki,Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	2 (sites)
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3 (sites)
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M., Sumi.N., Ishii.Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsunoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Matshiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawat,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.
TITLE	Riken integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	4 (sites)
AUTHORS	The RIKEN genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
MEDLINE	5 (bases 1 to 1507)
PUBMED	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bull,C., Carninci,P., Fukuda,S., Fukushima,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hirooka,T., Horii,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,O., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kutihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numata,K., Ono,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schirral,L., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,Y., Tejima,Y., Toyota,T., Yamamura,T., Yamanka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki. The Institute of Physical and Chemical Research (RIKEN), Laboratory for genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	Please visit our web site ( <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> ) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAATTCCTCAGTAAATTAAATTAATCCCGCCCCCCC 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapted of sequence [5' GAGAGAGAGATTCCTCAGTAAATTAAATTAATCCCGCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.
FEATURES	Location/Qualifiers 1..1507 /organism="Mus musculus"
SOURCE	

Query Match	Best Local Similarity	Score	DB	Length
Matches 1092; Conservative	78.5%;	Pred. No. 6.3e-146;	Mismatches 206; Indels 93; Gaps 1;	
Query	1670	tgtagccgacagcgtcgaattcgaagcgcgcgcagggctcgtgcgcaactgtgtgc	1729	
Db	1	TCGACCCACACAGCCGCAATTCAGAGTGGCGAGGGCTGTGGCGCAACGTGGCTGC	60	
Query	1730	gagtgcaatgacatgacagggagagcccaagcgaagctctgcggtatccttgcaagt	1789	
Db	61	GGATCCAGTACATGGCGCGGTGAAGCAAGCGCAGGCTTGGCGGCACTTTGGGAAT	120	
Query	1790	cgagctgcagbaattaccgcgcgcgcgccttacaccggtgtgttaccatacaagctcc	1849	
Db	121	CTAGCTGACGCGAATTCACCGAGAGGCTTTACACCACTGCTTATCAACAACTCTC	180	
Query	1850	ccgagcttcaagagagttcaagctgcacatcttcacgcgcgtcgtgcagagaaacatacc	1909	
Db	181	CGAATTTACAGAGGAATTCAGAGCTACGACTTCCTCGTGGGAGACCGAATCATCAC	240	
Query	1910	tgctgttcaacttcaacatgtcaagctgcagcccgcgccggcgcaactgccttggagaac	1969	
Db	241	TCCTTTTACCTTACCACTGTCAGCTGCACGCCCGGCGAGGAACAGCCTTGGAGACGC	300	
Query	1970	ccgttgagcttacttgatcccaactgcgtgcagcaagggcgcttgagagacggcccttct	2029	
Db	301	CGTGGGCTTCACCTTGGATCCCTCTGTTCACAACTGCGCGCTGAGAGATGGTCCCTTC	360	
Query	2030	gtctcccaatgtctgtgagacagccgcgcgcgcgaactatctcgtgtctcaacccgaatgtg	2089	
Db	361	GGCTGCGCCGTGCCGAGACCAAGCCCTCCACCACTGCTGCTGACGCCAGATGTAG	420	
Query	2090	cgcttcgcagcatgcgctgtgtgagacgtgtcaaaagggcglttcaagltgtgagctcaag	2149	
Db	421	CGCTGCTGCATGCGCTGGTGGAGCGCCACAAAGGCGTGTTCATGTTGGAGCTCACCG	480	
Query	2150	ccgtgtcctctgtgcaaccccccaaggaaccccttaccttgagaacaattctcaaccctggtgacag	2209	
Db	481	CGGTGCTGCCTGTGACACCGCGAGAGCCCCACTTGGATTAAGTTCCTTCACAGCTGTGATCAATG	540	
Query	2210	tcccttgagagagagccttcacatccgcgtcaagaaacactgtgtgagcgaagggcaag	2269	
Db	541	TTCTAGAGGAAGGATCTTCCGCTTCACACTAAGAGAGCACTGCTGAGACTTAAGGCACTA	600	
Query	2270	tggaagcagagagctgtgcggcgcaagltcttgcaagcaactgcgccttgccagccccgaacccttg	2329	
Db	601	TCGAGAGAGAGTTGGAGAGCCAGCTGGGCGACGCTTGCGTGCACAGCCAGAGCCCTAG	660	
Query	2330	tgagccttctccacacagctgtgtgacaagctcgtgtgtgtgtatcaagcccccatga	2389	

Db	661	TACCCCTTTTCCACCCTCGTGTGCTAGACAAGACTTTCGCCGTTGGTTGTGCGGCAACCCATCA	720
QY	2390	tcagtgagcagaatgtgtgaaccttgagccgttgagacctttgaagcaaatgccccatgtatga	2449
Db	721	TTTTGTGTGCAGAGTGTGAACCTGGGTCTGAGAGGGGCTTTGAAGCTATGCTCATGTAGCA	780
QY	2450	gaccttgcacccgagacccctgagagagagcccaagatgtcccgcggtcaactgtcccaagctg	2509
Db	781	GCCTTTGTGCACCGGAACCTGGAGGCTGTGCACAAGACTCCGTGGCCACTCCCACTGCTGG	840
QY	2510	ctgcctacgttccactacctaagcctttcgccttcccttccgtacgtatgaccaccaaccctggatggg	2569
Db	841	CCCTCATGTGCTCCACTAGCCGCTTTCGCGTCCTCGGTGTGTGATCCTACACCTGCCAGGGGAAG	900
QY	2570	ccccctcaatgtgaagagcagagctgtcccaacacttggcccggtgtctctgtgtcgccccgaagcc	2629
Db	901	CCCCCTCACCAACACTGTGTCAAGGCTGGCCACACTGGCCCGCTGTGTGCTCCCGCCGACAGCC	960
QY	2630	tctaccttgccggtttccaaagagcatcagcagcagcaaccctgtacccttcgcgttggccccctg	2689
Db	961	TGTACTGGCAGCGGTCTAAGAGCATCAGCAGCAGCAACCTTACCTGCGCTGTGTGCTCCTG	1020
QY	2690	gtctctgtgataagagaggtttcccgcatcccttggccaagc-----	2727
Db	1021	GCCTGTGTGACGATGAGTGTGCTCCGCTCTTACGCGACGAGGCGCATTCAGCCCAACTCTA	1080
QY	2728	-----	2727
Db	1081	GCCGGGCGCTCCTTATACCTCTGAGGCTCATCTCGGCGCTTGGCAGCGCCCTCACTGAGAC	1140
QY	2728	-----aagctgtcttcacgagagctgtgtctgtcagtggtgtgttcacgacagatg	2776
Db	1141	ACACTGTGCAGAGAGCTGCTGCACGAGAGACTGTGCTGTGCACTGGGAGGAGTGCAGCGGAGTG	1200
QY	2777	ccgtacgcgagagccatccctccagagcagcgcgtgtcttcttcacgtcatatgtgtgaagatga	2836
Db	1201	CGGTTGCTAGCTAGTCTGCTGCGACATGCTGTGTTCTTCTTCCAGCTCACTAGTAAAGCA	1260
QY	2837	tggcgctgacactgtctgtcttggccagagcagctaaagacaaacccgcaagctgtgccttcccg	2896
Db	1261	TGGAGCTTCATGTGCTTCTGGGCGCAGCGCACTGGACACTCCCGCAACGTCGCTTCCCTG	1320
QY	2897	gagcgcttccctggagacacatcactgtctgtgtgtgtctgttggcctgtgagatgcataccccc	2956
Db	1321	GGCGCTCTCTGATGATGACTTGCGCCCTGTGTGCTTCGTGGGCGCTGGAAATCATCACC	1380
QY	2957	gtgtccacaag 2967	
Db	1381	GAGTCCATTAAG 1391	
RESULT 4			
BM473590/c			
LOCUS	BM473590	1112 bp	mRNA linear EST 05-FEB-2002
DEFINITION	AGENCOURT.6466302 NIH_MGC_88 Homo sapiens CDNA clone IMAGE:5562000		
ACCESSION	BM473590		
VERSION	5', mRNA sequence.		
KEYWORDS	BM473590.1 GI:18522632		
SOURCE	EST.		
ORGANISM	human.		
	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
REFERENCE	1 (bases 1 to 1112)		
AUTHORS	NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> .		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a>		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Agencourt Bioscience Corporation		

clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: LLAM12290 row: 1 column: 01  
 High quality sequence stop: 559.

FEATURES	Location/Qualifiers
source	1. .1112

```

/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="IMAGE:5562000"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dt primed. Average insert size 1,767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."

```

Query Match	12.5%;	Score 797;	DB 10;	Length 1112;
Best Local Similarity	96.8%;	Pred. No. 1.3e-133;		
Matches 867; Conservative	0;	Mismatches 21;	Indels 8;	Gaps 5

QY	5484	tggaagccgctacttgatccct-----acagagctaaagacccgggtgacctcttga-c	5538
Db	916	tggaaacacccgctacttcttatttacttgcacacactccaagaccggggacttactttgacc	857
QY	5539	cgcaactatgagc-ttcgcacatctctgtctcgaacgcggttcacg-ccgagatggacg	5598
Db	856	cgcaactatggggcttctggcactttccgtgtttggcaccgcggttcacgncggatggggcg	797
QY	5597	cacacggggagcgtcgcccgagcaacaaagcgttaagacgctgtctcagcaacgcacgcct	5656
Db	796	cacacggggagcgtcgcccgagcaacaaagcgttaagacgctgtctcagcaacgcacgcct	737
QY	5657	tcgcc-tcacatcaagactgcacatccgttgtgtgcacacgggaggaagcgtgtctgcagca	5715
Db	736	tcgccnctacatcaagactgcacatccgttgtgtgtgcacacgggagagacgcgtgtgcagca	677
QY	5716	gtgtaggttgccatctgagagacatgcagaaagaagacacgggagcctgctcttgccagag	5775
Db	676	gttgagggtgcccactgcagggacatgcagaaagaaacacgggagccttgctcttcacacag	617
QY	5776	cagagaccacacagatgctaaagatgctacagatgcttcacaggtctctgtaaagccac	5835
Db	616	cagagaccacacagatgctaaagatgctacagatgcttcacaggtctctgtaaagccac	557
QY	5836	gtgaacacaggctccctctgtaggttgccacagtgcttttaagcagatcccgaaagacc	5895
Db	556	gtgaaacacagggtccctctgtaggttgccacagtgcttttaagcagatcccgaaacccc	497
QY	5896	aagctctcggagatatacaacaatattggcgtctctgaagaagatttgtgaagaatgt	5955
Db	496	aagctctcggagatatacaacaatattggcgtctctgaagaagatttgtgaagaatgt	437
QY	5956	gagagatgcctgtcgcgaataaaataagagccctgatttggccgagacaaagagataccacgct	6015
Db	436	gagagatgcctgtcgcgaataaaataagagccctgatttggccgagacaaagagataccacgct	377
QY	6016	gagcttgtagagcgcaactactgcgcgctgcgcggagagctctgcagccctcttaaccaagc	6075
Db	376	gagcttgtagagcgcaactactgcgcgctgcgcggagagctctgcagccctcttaaccaagc	317
QY	6076	ctgcgccagctgaatggacacccacccaccccgagccccaagaatctctgaacagagcaat	6135
Db	316	ctgcgccagctgaatggacacccacccaccccgagccccaagaatctctgaacagagcaat	257
QY	6136	tctccgaagtcagacctctgagcccaagaagaccaaagctgtacctagaggaaacacagac	6195
Db	256	tctccgaagtcagacctctgagcccaagaagaccaaagctgtacctagaggaaacacagac	197

QY	6196	ccggagcttcgaagctgctgtgctgcggaaggggagatcggccctggtggcccaatggcctgtggg	6255
Db	196	ccggagcttcgaagctgctgtgctgcggaagggagatcggccctggtggcccaatggcctgtggg	137
QY	6256	gtgacacacacttacttgggagctggccctcgtcccccgtgtgcccatcgtgtgctacgt	6315
Db	136	gtgacacacacacttacttgggagctggccctcgtcccccgtgtgcccatcgtgtgctacgt	77
QY	6316	atgcttcctccctcttttaatttaaatggttttataagcaaaaaaaaaaaaaa	6371
Db	76	atgcttcctccctcttttttaatttaaatggttttataagcaaaaaaaaaaaaaa	21
RESULT	5		
LOCUS	AK004816	2694 bp mRNA linear HTC 19-JAN-2002	
DEFINITION		Mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200017A24;homolog to BAI65P24.1.1 (NOVEL PROTEIN SIMILAR TO DROSOPHILA CG6630 AND CG1376, KIAA1050, RAT TRG (ISOFORM 1)) (FRAGMENT), full insert sequence.	
ACCESSION	AK004816		
VERSION	AK004816.1	GI:12836282	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (strain: C57BL/6J) adult male lung cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library clone:1200017A24.		
ORGANISM	Mus musculus		
REFERENCE	Enkairi, Y., Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (sites)		
TITLE	Carninci, P. and Hayashizaki, Y.		
JOURNAL	High-efficiency full-length cDNA cloning		
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)		
PUBMED	99279293		
REFERENCE	10349636		
AUTHORS	2 (sites)		
TITLE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
MEDLINE	Genome Res. 10 (10), 1617-1630 (2000)		
PUBMED	20499374		
REFERENCE	11042159		
AUTHORS	3 (sites)		
TITLE	Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, U., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasai, Y., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsui, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, T., Kira, A. and Hayashizaki, Y.		
JOURNAL	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipillary sequencer		
MEDLINE	Genome Res. 10 (11), 1757-1771 (2000)		
PUBMED	20530913		
REFERENCE	11076861		
AUTHORS	4 (sites)		
TITLE	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.		
JOURNAL	Functional annotation of a full-length mouse cDNA collection		
REFERENCE	Nature 409, 685-690 (2001)		
AUTHORS	5 (bases 1 to 2694)		
TITLE	Adachi, T., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bonio, H., Brownstein, M., Bull, C., Carinci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numata, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schirml, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T.,		

TITLE  
JOURNAL

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer 15' GAGAGAGAAGCGCCGCCACTGGATGGATTTTCTTTTTTTTN 3', cDNA was prepared by using trihalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence15' GAGAGAGAAGACCAAGCAGCTCAATTAAATTAATAAOCCTCCCCCCC 3'. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI, 3' end: XhoI. Host: SOLR.

FEATURES  
SOURCE  
Location/Organism  
1..2694  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="MGD:MGI:1907255"  
/db\_xref="taxon:10090"  
/clone="1200017A24"  
/sex="male"  
/tissue\_type="Lung"  
/clone\_id="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
22..1563  
/note="data source:SPTF, source key:09H1Q3, evidence:ISS homolog to BA165F24.1.1 (NOVEL PROTEIN SIMILAR TO DROSOPHILA CG6630 AND CG11376, KIAA1058, RAT TRG (ISOFORM 1)) (FRAGMENT)  
putative"  
/codon\_start=1  
/protein\_id="BAB23567.1"  
/db\_xref="GI:12836283"  
/translation="MQTPPPMVEELCLNLSLYTVKAREFOEDPEMLMDLMYRI AKSQASPDRLTWLOMAKNHTKKCFTEAAKGLVLAALVAEYSLEMDHSYLPG SVSNQNSNVLTSESAVSODTLSPDEDCSGRGFTSSGIVGLEOAEFLPSGVGLYE TVNBYKYLVPILEAHNDPRLKSTSHDKLRADNTINKKHRRFGTYPRVGYSGF GDUDEOFVKREPATITKLPEISHLBSFYQCFSGAEEVIKISTDPVTKRLDPNNAY IITFEVPFDEEMKRDVTYFEKNFLRFMTTPPLEIKIGSDLEADHRNRVYLT TMHAFVFKTRIRVSKSEEFVLPIETVALIDMKKTQLVLAVTHOEPRDAKMOLVQ GSMAFTVNQSLDEAVQVLAIEIPADPKLYRHNNKLRLCFEPAINTCGEAVEKRRLITT AEOREYQDELKNNKRLDSLRPIEKIPELVKPIPRVDQSQRDSHFHSFKCEIQ LSGCS"

polyA\_signal  
2677..2682  
/note="putative"  
2694  
/note="putative"  
BASE COUNT  
752 a 646 c 682 g 614 t  
ORIGIN

Query Match 12.1%; Score 768.4; DB 11; Length 2694;  
Best Local Similarity 70.7%; Pred. No. 1.9e-128;  
Matches 1040; Conservative 0; Mismatches 426; Indels 6; Gaps 1;

OY 4615 tatctgtagggagcattgcgtcggaacgcccttcggaagccaggttcaggactg 4674  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 1 TTTCACGAGAGGACGCGGCATCGAGCAACTCTCTTTTCCCATGCGAGTGAGGAACTT 60  
OY 4675 atgttcaaccctgcacatcatctgaagcaacccggttgaaatcgaaggaacaccagaagac 4734



```

Db      61 CTGCGCATCTGAACAGCATTTCTGTACGACACAGTAGATGAGGGAATTCAGGAAGAC 120
Qy      4735 ccgagagtgctcgcagccatgacgaatgtcccgaggctcaccagggctcaccggac 4794
Db      121 CCGAGATGCTTATGAGCCCTATGATGACATTAATGCAAGAGCTACCAAGAGCATGCTGTAC 180
Qy      4795 cctcgctgcagccgtgtgtgcgaacatgagccgggaagcagcgaggtgtggcaaccagcc 4854
Db      181 CTTGGGCTGACTTGGCTCCAGAAATGCGCAGAAACACACTAAGAAAGTCTCTTCA 240
Qy      4855 gaagcgccagtgatgtgacgagcgagcgccctcgtgtgagtaactcgcctgtc 4914
Db      241 GAGGCCCCCATGTGCTGGGTGCAATGCAAGCGCCCTGTGCTGACGATCTGAGCATGCTG 300
Qy      4915 gaagcagcgagccacccgtggcgtgtgtgttctcttcgaagaactcgaatccacgtg 4974
Db      301 GAGGACCAACAGTACTGCGGGTGGGAGCGTACGCTTTCAGAAATTTCTTCCAAATGTG 360
Qy      4975 ctgagagagtcgcacatctccgcagacatctgttcgcccgcagagagaggtctctgc 5034
Db      361 CTTGAGAGAGCTGTGAGTCTGTGATGACACTGTGTCACTGATGAGGAGCGGCGTATGCTCT 420
Qy      5035 gggaagcacttcacatgagctggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5094
Db      421 GGTGCTACTTCACTGAGAGTGGCTGTGGGCTCTCTGAGCAGGCTGTGGAGCTCTTC 480
Qy      5095 accatggcgagctcgcagcgagtgagtgatgtgtacagaagactccatccatccg 5154
Db      481 AGCAGCGGAGGCTTGTGACGAGCGTTAATGAAGTCTACAAAGCTGGTCACTCCATTCCTG 540
Qy      5155 gaagccacgcgtactacaagaagctgtgcgagtgcaagcgcaactgtcagaagagctc 5214
Db      541 GAGGACACAGAGATTTCCGGAAGCTGACCTCCACTCACACAACTGCAGAGAGCGCTTC 600
Qy      5215 accaagatctgacacagatctccgctgtgagcgctgtgtgtgtgtgtgtgtgtgtgtgtgt 5274
Db      601 GATTAACATCAACAAAG-----GACCATTAAGAGATGTTTGGAGCTACTTCCAGATT 654
Qy      5275 ggtctcaagcgccacatctcggtgacgtgacgagcaagagttgtgtacaagagcgca 5334
Db      655 GGTTCACGAGATCCCATTTGGGAGATTGTGATGACGAGAGTTCGTATCAAGAGAACCC 714
Qy      5335 tgcatacgaagctgtgcagagatctcaacacgcgtgtgagagttctacaagagagatt 5394
Db      715 GCATACGAGAGCTCCCGGAGATCTCATAGACGAGGATTATGCGCCAGTGTTC 774
Qy      5395 ggagcagagctgtgtgagatctcaagaactctaccctgtgtgacgaagtcacagctgtgac 5454
Db      775 GGTGCAGAGTTTGTGGAAGTGAATGAAGACTTCACTCACTGAGCAAAACCAAGTTGAT 834
Qy      5455 tcaaaaagagctacatccagatccagatcagatgtgtgaacggtacttatacctcagagctc 5514
Db      835 CCTTAACAGAGCTTAATCAATCAATGATCATTGTTGTGAGACCTTACTTGTATGAATGTGAGATG 894
Qy      5515 aagagccgggtgaactactgtgaacgaactatgtggctgtgcacattcctgtctgcag 5574
Db      895 AAAGAGCGGGGTGACTTCTCGAAGAATTTCAACTCCGAGGTTTCATGTACACCACC 954
Qy      5575 ccgttaccgctggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5634
Db      955 CCGTTCACCTGTGAGGGGAGACCCCGGGGAGACTTTCATGAGCAACACCGCAAAACACC 1014
Qy      5635 ctgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 5694
Db      1015 GTGCTCACACACATGACAGCGCTTCCCTTACATTAAGACCAAGATCCGAGTCAAGCCAGAA 1074
Qy      5695 gaagagacggtgtgtgacgagcagtgagagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5754
Db      1075 GAGAGAGTTCGTTTGAAGTCCGATTTGAAGGTTGCCATTGAAGATGAAGAAAGAACCTGTG 1134
Qy      5755 gaagtgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5814

```

```

Db      1135 CAGTTAGCCGTGCGCCTACCTACACAGAGCCCTCGATGCAAAAGATCTGCAAAATGTACTG 1194
Qy      5815 cagggctctgttagagccaccagctgaaccaggggtccctgtgaggtgtgtgtgtgtgtgtgtgt 5874
Db      1195 CAGGCTCTGTAGAGCCACTGTAAATCAGGACACTGGAGGTGTGGCCCAAGTGTCTTG 1254
Qy      5875 gcaagatcccggaagaccccaagctctccggatcacaacaatgtgtgtgtgtgtgtgtgtgtgt 5934
Db      1255 GGTGAATTCAGCTGACCCCAAGCTCTACGACATCAACACAGCTGAGGTTGTGCTTC 1314
Qy      5935 aagagctctgcaagaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5994
Db      1315 AAGGAGTTCAATATTCGATGTGGAGAGGCCGTGAGAGAAACAGCGCACTATCACCGCA 1374
Qy      5995 gaccagagaggtatcacaccggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 6054
Db      1375 GAGCAGCGGAGACTACACAGAGAGTGAAGAAAGACTCAACACAGCTGTAGAGACAGCTC 1434
Qy      6055 cagccctgtctaccagcgctgtcccaagct 6086
Db      1435 AGGCCCATGATTGAGCGGAAATCCAGAGCT 1466

```

```

RESULT 6
BI770146 796 bp mRNA linear EST 25-SEP-2001
LOCUS 603053332P1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5203063 5'
DEFINITION mRNA sequence.
ACCESSION BI770146
VERSION BI770146.1 GI:15761724
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

```

```

REFERENCE 1 (bases 1 to 796)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LNA11508 row: p column: 08
High quality sequence stop: 793.

```

#### FEATURES

```

source
1..796
location/Qualifiers
1..796
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5203063"
/clone_id="NIH_MGC_122"
/lab_host="DH10B"
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
Site:1: NotI; Site:2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 026. Note:
this is a NIH-MGC Library."
BASE COUNT 188 a 228 c 231 g 148 t 1 others
ORIGIN

```

```

Query Match 12.0%; Score 762.4; DB 10; Length 796;
Best Local Similarity 99.6%; Pred. No. 2.2e-127;

```

	Matches	785;	Conservative	0;	Mismatches	1;	Indels	2;	Gaps	2;
Oy	5275	ggcttctacgcgcgcaccattcgttgacccttgatgtagcagagatttgltgtacaaggagcca	5334							
Db	1	ggcttctacgcgcgcaccattcgttgacccttgatgtagcagagatttgltgtacaaggagcca	60							
Oy	5335	tcgatcacgaagctgtgcagagatctcacaccgcgtgtaggagattctacaaggagattt	5394							
Db	61	tgcgacacgaagcttgccagagatctcacaccgcgtgtaggagattctacaaggagattt	120							
Oy	5395	ggcgagcgcgcgcctgttgatattatacaaaactcttaccctgttgagaaagtgcaagtttac	5454							
Db	121	ggcgagcgcgcgcctgttgatattatacaaaactcttaccctgttgagaaagtgcaagtttac	180							
Oy	5455	tcacaaagaagccctacatccagatcacatcgtatgtgagaccgtactttgatccacagctc	5514							
Db	181	tcacaaagaagccctacatccagatcacatcgtatgtgagaccgtactttgatccacagctc	240							
Oy	5515	aagagccgggtgtgacctactcttgaccgcgaactatgggtcttcgcacattccctgttcgacg	5574							
Db	241	aagagccgggtgtgacctactcttgaccgcgaactatgggtcttcgcacattccctgttcgacg	300							
Oy	5575	ccgttcaacgcgcgataggggcgcgcacacggggagggcgcccgccgagacaacaagaagtgtaagc	5634							
Db	301	ccgttcaacgcgcgataggggcgcgcacacggggagggcgcccgccgagacaacaagaagtgtaagc	360							
Oy	5635	ctgctcagcacacgcgaaccagccttccctctacatcacaagaactcgcatacgtgtgtgtgcacgg	5694							
Db	361	ctgctcagcacacgcgaaccagccttccctctacatcacaagaactcgcatacgtgtgtgtgcacgg	420							
Oy	5695	gaggaagaacggtgtgtgagcgcagtgtagagtgtgcacatcgagagacatgcagaagaagacacgg	5754							
Db	421	gaggaagaacggtgtgtgagcgcagtgtagagtgtgcacatcgagagacatgcagaagaagacacgg	480							
Oy	5755	gagctgagcctctgtgcaccgcgagagaccaccaacagatgtcttaagatgtcaagaatgtgtgctt	5814							
Db	481	gagctgagcctctgtgtgaccgcgagagaccaccaacagatgtcttaagatgtcaagaatgtgtgctt	540							
Oy	5815	caaggctctgtgtagggccacacgttgacaagaaggttccctctgagggttgccagagtgttttta	5874							
Db	541	caaggctctgtgtagggccacacgttgacaagaaggttccctctgagggttgccagagtgttttta	600							
Oy	5875	gcagagatcccgjaaagcccagaactcttcctgggcatcacacaatttcgctctgtctc	5934							
Db	601	gcagagatcccgjaaagcccagaactcttcctgggcatcacacaatttcgctctgtctc	660							
Oy	5935	aaggactcttgtaaaagaatagtgtgagatgtgcctgcgggaaa-aataagggccctatttgggcc	5993							
Db	661	aaggactcttgtaaaagaatagtgtgagatgtgcctgcgggaaa-aataagggccctatttgggcc	720							
Oy	5994	ggaacacagaagajtaacacacgcgtgagctgtgagtgcaactac-tgcgcgcctgcggagagctc	6052							
Db	721	ggaacacagaagajtaacacacgcgtgagctgtgagtgcaactac-tgcgcgcctgcggagagctc	780							
Oy	6053	tgcagccc 6060								
Db	781	tgcagccc 788								
RESULT	7									
LOCUS	B1831979/c	815 bp	mRNA	linear	EST 04-OCT-2003					
DEFINITION	603076025P1 NIH_MGC_119	Homo sapiens	CDNA	clone IMAGE:5167759	5',					
ACCESSION	B1831979									
VERSION	B1831979.1	GI:15943529								
KEYWORDS	EST.									
SOURCE	human.									
ORGANISM	Homo sapiens									
REFERENCE	1	(bases 1 to 815)								
AUTHORS	NIH-MGC	http://mgc.ncl.nih.gov/.								

FEATURES	source
JOURNAL COMMENT	<p>National Institutes of Health, Mammalian Gene Collection (MGC)  Unpublished (1999)  Contact: Robert Strausberg, Ph.D.  Email: cga@bbs-remail.nih.gov  Tissue Procurement: Life Technologies, Inc.  cDNA Library Preparation: Life Technologies, Inc.  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  DNA Sequencing by: Incyte Genomics, Inc.  Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at:  http://image.lnl.gov  plate: LMNL1417 row: a column: 08  High quality sequence start: 25  High quality sequence stop: 812.  Location/Qualifiers</p>
BASE COUNT	<p>140 a 239 c 257 g 177 t 2 others</p>
ORIGIN	<p>1. 815  /organism="Homo sapiens"  /db_xref="taxon:9606"  /clone_image="5167759"  /clone_lib="NIH_MGC_119"  /tissue_type="medulla"  /lab_host="DH10B"  /note="Organ: brain; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note this is a NIH_MGC Library."</p>
Query Match	11.9%; Score 758.4; DB 10; Length 815;
Best Local Similarity	99.6%; Pred. No. 1.2e-126;
Matches 770; Conservative	0; Mismatches 2; Indels 1; Gaps 1
5533	tttgacgcgaactatggtgtcgacattccctgttcgacgcgcgttcacgcgcgaatgg 5592
803	ttttgacgcgaactatggtgttcgacattccctgttcgacgcgcgttcacgcgcgaatgg 744
5593	cgcgcacacgaggagctgcccgcgagcaacaacgctgaagacgctgctcagacacgaac 5652
743	ccgcgcacacgaggagctgcccgcgagcaacaacgctgaagacgctgctcagacacgaac 684
5653	gccttccctaca-tcaagactcgcatccgtgtgtgcccacgcgggagagagcgtgtctgac 5711
683	gccttccctacaagctcaacgctcgcgtcgtgtgtgcccacgcgggagagagcgtgtctgac 624
5712	gccagtgaggagtgcccatgaggacatgagaagaagaacagcggagctgacctgtccac 5771
623	gccagtgaggagtgcccatgaggacatgagaagaagaacagcggagctgacctgtccac 564
5772	cgaagcagacccacacagatgcttaagaatgcttaagaatgcttcaagagctctgttagggcc 5831
563	ccagcagcagacccacacagatgcttaagaatgcttcaagagctctgttagggcc 504
5832	caacgtgaacacaggtgcccttgagggtgcccagatgtttttaagcagagatcccggaaga 5891
503	cacccggaacacaggtgcccttgagggtgcccagatgtttttaagcagagatcccggaaga 444
5892	cccccaagccttcgcgcacacacaacaaattggcgtctctctcaagagatcttgcgaaga 5951
443	cccccaagccttcgcgcacacacaacaaattggcgtctctctcaagagatcttgcgaaga 384
5952	atctgagagatgctgcttcggaataaataagcctctgattggccggaacagaagagtacaa 6011
383	atgtgagagatgctgcttcggaataaataagcctctgattggccggaacagaagagtacaa 324
6012	ccgtgagctgagcgcgaactactgccgctcgaggagagctctgcagccctcttaacca 6071

Db 323 CCGTAGAGCTGGAGCGCACTACTGCGCGCTGGGGAGGCTGTGACAGCCCTGCTTACCCA 264

Qy 6072 ggcgcgtgccccagctgatalgagccaccaccgagcgcctcgaagaaactccttgaacagagc 6131

Db 263 GCGGCTTGCCCGACTGATGGCACCCACCCGCGCTCAGGAATCCTTGAACAGAGC 204

Qy 6132 aagttccgaagagcagacctgtagcccaagaagcaaaagcttacctagaaggaacca 6191

Db 203 AAGTTCCGAAAGCAGAGACTGTGAGCCACACAAGAGCAAAAGCTGTACTAGAGMAACA 144

Qy 6192 ggaacccggagcctgaagctgtctgtcgcgaaggagagtcgcctgtgtgcacactgtgcgc 6251

Db 143 GCACCGGCGCTCAGCTGTCTGTGTCGCGAGGAGCTGTCACTGGTGTCCCTGCGCTG 84

Qy 6252 tgggggtgacacacactgtactgtggtggtgcctcgcctcgtgtgtcccatc 6304

Db 83 TGGGGTGACACACTGTACTGTGGGCTGGGCCCTCTGCCCTGTGTGCCCATC 31

RESULT 8

LOCUS B1820302 927 bp mRNA linear EST 04-OCT-2001

DEFINITION 603036864F1 NIH\_MGC\_115 Homo sapiens cDNA clone IMAGE:5178087 5',

ACCESSION B1820302

VERSION B1820302.1 GI:15931852

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 927)

AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D. Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov) Tissue Procurement: Life Technologies, Inc. cDNA Library Preparation: Life Technologies, Inc. DNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL) Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at: <http://image.llnl.gov> Plate: L14M1143 row: 0 column: 16 High quality sequence stop: 806.

FEATURES

source

1..927

Location/Qualifiers

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5178087"

/clone\_lib="NIH\_MGC\_115"

/lab\_host="DH10B"

/note="Organ: pooled brain, lung, testis. Vector: pCMV-SPORT6. site\_1: NotI; site\_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH\_MGC Library."

BASE COUNT 214 a 257 c 292 g 164 t

ORIGIN

Query Match 11.8%; Score 751.4; DB 10; Length 927;

Best Local Similarity 97.9%; Pred. No. 2.2e-125;

Matches 825; Conservative 0; Mismatches 11; Indels 7; Gaps 6;

Qy 3906 ggggaaagagccttgacgcatacaagcctcatcctcaaaatctcggatataa 3965

|||||

Db 1 GGGGAAAAAGCCTTTGAAGCATCACAGCCTCATATTCAAAAAATTTGTGATATGAA 60

Qy 3966 ggcgcgtgtagaagaagcattctgggtacatcggagctcgacaagaatggttggcg 4025

Db 61 GCGGCGCTTAAGAGAGCATTCTGGTACCATGGAACTGGACAAGAAATGTTCCGCG 120

Qy 4026 aagtcgtagaagagcccgcttctgggaatccagaaacgtgtcgtcgtggaagagcgtcac 4085

Db 121 AAGTGTGAGAGAGGCCCG-TTGGGAATCCGGAGAATGTGGCTGGGGGAAGAGCGTAC 179

Qy 4086 acacttggaagcaaaccttaagacgcgtgtgagacaagaaccaagatgaatgaacagagc 4145

Db 180 AACATGGAGGAAACCTTCAGACCCGCTGAGACAAACCAAGCATGAATGGAACACGAGGC 239

Qy 4146 ctgtgtgaagggaaacctgagcaaccggaggaagcctagtgttcttgagacacatggagat 4205

Db 240 CTGTGTGAAGGGAACCTTGGCAACCGAGGCAAGCTTATGTGTTTGGACACACTGGAGAT 239

Qy 4206 catcgtgacagcgtgtgatgtcttcagaagcccggaagacgctcttggggcagtgctgaa 4265

Db 300 CATCGTCAGAGCGGATGATCTTTCAGAAAGCCGGAGAGCGTCTTGGGGGCAAGTGTGAA 359

Qy 4266 ggtgtgtcgtacagcgtgtgagcagtgccagagtgccctctcttcagacatggcgtgac 4325

Db 360 GGTGTGCTGTACAGCCTGGGCACTGGCCAGAGTGCCCTCTTCTTGGACGATGGCTGGC 419

Qy 4326 caccagaagggcccttgtgtccaaagttccggagctgtcttggaggaagcaagagct 4385

Db 420 CACCCAGAGGCGCCTTGTGTCAAGTTCCGAGACTCTGTGTCAAGGAGACAGGAGCT 479

Qy 4386 gtgtgcccacctgtgtcctgaagcctcctcaagacaactgtgtgcagccgcatcagacacatcg 4445

Db 480 GTGTGCGGACCTGTGTGCGTCAAGGCTCTCAACGACATGTGTGGAGCGGATCAACGACATCCG 539

Qy 4446 caagcaagcagcgcctcgtgtactcgtgtcctcaatgtgcacagaactcgtgagatggccaac 4505

Db 540 CAGCCAGCCCGACCGCTCGCTGCTACCTGCTCATCGACAAACTTCAGATCCGCCACAA 599

Qy 4506 ctgtgcccgtgtgaagatgacaggtcaccc-atgtctctcgtccctgtgt-ggggagcagc 4563

Db 600 CTTTGCCCGTGTGAAGATGCAAGTGCACCATGTCTCTCTGCTCCGTGGGGAGCAGC 659

Qy 4564 cagaactcagtgaaagagcagctgtgcagcttcaatc-aaaacatcctcaactatgtcga 4622

Db 660 CAGAACTTCAATGAAGAGCACTCGACGTTCATCAAAAAACATCCTCACTATGCTGA 719

Qy 4623 ggaagga-catgggctcgcggagacagcacttcgagagcaggtgccagagactatgttca 4681

Db 720 GGAGGACCATGGGCTCGGGACAGCACTTCGCGAGAGCAAGTCCAGACCTGATGTTC 779

Qy 4682 acctgcacatgatcctgaagcagacaggtgaagat--gaaagaaacacaggaagacacctga 4739

Db 780 ACCGCGCAATATCTCTACGAGACAGCGSTGAAGATTCAGAGGGGCCCCAGAGGAGACCTGGA 839

Qy 4740 gat 4742

Db 840 GAT 842

RESULT 9

LOCUS B1830310 770 bp mRNA linear EST 04-OCT-2001

DEFINITION 603073042F1 NIH\_MGC\_119 Homo sapiens cDNA clone IMAGE:5164762 5',

ACCESSION B1830310

VERSION B1830310.1 GI:15941860

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 770)

AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov  
Plate: L1A11409 row: d column: 11  
High quality sequence stop: 760.  
Location/Qualifiers  
1. .770  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5164762"  
/clone\_lib="NIH MGC.119"  
/tissue\_type="medulla"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dt primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH MGC Library."  
BASE COUNT 164 a 239 c 223 g 144 t  
ORIGIN

Query Match 11.6%; Score 738; DB 10; Length 770;  
Best Local Similarity 99.1%; Pred. No. 5.6e-123;  
Matches 763; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

241 gaattcccaagctgtagtctgtctgtcagcccggaatgcgcgacacagag 300  
|||||  
1 GAATTCCTCCAGCGATGACTGTGAGCTGCTGTCAGCCCGGGAATGCCGACGAG 60  
301 ccgggagatcccaagagtagaaactgtagtgcaggtgagggcggtgagatgat 360  
|||||  
61 CCCGGGATCCCPAGGATGAGAACTGAGTGCAGGtGAGGCGCGGTGAGATGTAT 120  
361 attggaactgggtatctgtccacagaagratcagtaactgagtgacacacagccc 420  
|||||  
121 ATTGAGGACTGGCTCATGTGTCACAGAAAGTATAGTACTGAGTGCAGATACAGCC 180  
421 gtccaccacagacacacagcgagagcgacagaaaggctcccccagagctttgagcag 480  
|||||  
181 GTTACCCACAGACACACAGCGGAGCGACAGAAAGGCTCTCCCGCCAGAGTCTTTGAGCAG 240  
481 gatgtctctggaagagagaggtccggccctgaagagactcgaatgactccgggctgctcg 540  
|||||  
241 GATGCTTCTGGAGACGAGAGNGTCCGGCTGAGGACTGGAATGACTCCCGGCTGCTG 300  
541 ggcctcccggaagacacccctcgaagcagtggtgctctagcatcttcgaactgagagac 600  
|||||  
301 GGCCTCCCGGAGACACCCCTCGAAGCAGTGTGCTCTAGATCTTGAGACTGAGAAC 360  
601 ctggagagctgactcatctgtccctctctgtctagagcgggcgcccgacgaagatgtgac 660  
|||||  
361 CTGGAGAGCTGACTCATGTGCTGCTCTGTGCTAGAGCGGGCGGCCCCAGAGATGTGAGC 420  
661 cgaggcaatgaaccccttcgaagcagcgaccccgccggccctctacacctacaccg 720  
|||||  
421 CGGGCAATGAACCCCTTGAGCGGACGCGGCCCGGCTCTACCCCTTACCCG 480  
721 gcaactgacgagagatgaagccgtgtaagcgtgtgacgcgacccagaccccgagagac 780  
|||||  
481 GCACCTGACGAGAGATGAAGCGGTGAGACGCTGTACGCCCTTAGACCCACCCCGCAGAC 540

QY 781 ttggacaagagatctgtgcaagtgtctgtcgtcaagltcgaagltgaattgacc 840  
|||||  
Db 541 TTGTGACAAAGAGATCTGTGTCAGTGTCTGCTCGCTCAAGTTGAGATTAATTTAGAGCC 600  
QY 841 atctttgagatctgtgtctgtatgtatgtcgggagagaaaaagaagatctccggaacttc 900  
|||||  
Db 601 ATCTTTGGGATCTTTGGCTCTGTATGATGTGCGGAGAAAAAGAAATCTTCGAAAACTTC 660  
QY 901 tacttgacctgaactcgaactcc-atgaagggtcgtcttcggcccaatgagccaccacc 959  
|||||  
Db 661 TACTTCGACTGGAATCTGGAGCTCCAATGAAGGGGCTGCTTCGGGCTCATAGGACCCACCC 720  
QY 960 tgcattccacccttgcgcctctg-ccatctctctgtgacctacc 1008  
|||||  
Db 721 TGCCATCTCCACCCCTGGCCCGCTCTGCGCCATCTTCTGTGAGATACCC 770

RESULT 10  
A1198543/c 794 bp mRNA linear EST 10-NOV-1998  
LOCUS A1198543  
DEFINITION g449f11.x1 Soares\_testis\_NHT Homo sapiens cDNA clone IMAGE:1753389  
3' similar to WP:P46H5.4 CE04590 RAT TRG GENE PRODUCT ;, mRNA  
sequence.  
ACCESSION A1198543  
VERSION A1198543.1 GI:3751149  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 794)  
NCT-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
Tumor Gene Index  
Unpublished (1997)  
CONTACT: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo  
, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCT-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.llnl.gov/db/ftp/image.html  
Insert Length: 856 Std Error: 0.00  
Seq primer: -40UP from G1bco  
High quality sequence stop: 444.  
Location/Qualifiers  
1. .794  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1753389"  
/clone\_lib="Soares\_testis\_NHT"  
/sex="male"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5' TGTTACCAATCAGTCACTGGAGCGGCGGCCCAATTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT733 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 149 a 219 c 236 g 190 t  
ORIGIN

Query Match 11.5%; Score 735; DB 9; Length 794;  
Best Local Similarity 97.7%; Pred. No. 1.9e-122;  
Matches 777; Conservative 0; Mismatches 15; Indels 3; Gaps 3;



|||||  
Db 541 CGGCTACTTACCATGGCGGCTTACGAGCGGTGATGAGTCTACAGAACTTCAT 600  
Qy 5145 ccccatccttgagagccacccgttgactaagaagctgagccggtgacaggaactgca 5204  
Db 601 CCCCATCTCGGAGGCCACCGCTGACTACAAAGACGTGGCCCGGTGACAGCAAACTGCA 660  
Qy 5205 ggaagccttaacccaagatcatgacacacagatctccggtgagagcggtgtcggagacgta 5264  
Db 661 GGAGGCGTTTCACCAAGATCATGACACAGAGTTCCGGCTG---GAGCCCGTGTGGGACGT 717  
Qy 5265 ttcccgctgaggtctctacagcgcccaacttcggtgacctggatgagcaggaagt 5318  
Db 718 ATTCCGCGTGGGCTTCTACGCGGCCCACTTCGTGTA-CTGGATGACAGAGATT 770  
RESULT 12  
LOCUS AM027453 793 bp mRNA linear EST 15-SEP-1999  
DEFINITION wt5a03.x1 NCI-CGAP GC6 Homo sapiens cDNA clone IMAGE:2515180 3'  
similar to WP:F46H5.4 CE04590 RAT TRG GENE PRODUCT ; mRNA  
sequence.  
ACCESSION AM027453  
VERSION AM027453.1 GI:5886209  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 793)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga03s-remail.nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/bbrp/image/image.html  
Seq primer: -40up from Glibco  
High quality sequence stop: 463.  
FEATURES  
source  
1. 793  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2515180"  
/clone\_lib="NCI-CGAP\_GC6"  
/tissue\_type="pooled germ cell tumors"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; Plasmid DNA  
from the normalized library NCI-CGAP GC4 was prepared, and  
ss circles were made in vitro. Following HAP purification,  
this DNA was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from a pool  
of 5,000 clones made from the same library (clonoids  
1257096-1258631, 1459064-1470983, and 1475592-1476743).  
Subtraction by Bento Soares and M. Fatima Bonaldo. "  
BASE COUNT 153 a 229 c 227 g 183 t 1 others  
ORIGIN  
Query Match 11.3%; Score 720.4; DB 9; Length 793;  
Best Local Similarity 94.7%; Pred. No. 8.3e-120;  
Matches 745; Conservative 0; Mismatches 42; Indels 0; Gaps 0;  
Qy 5572 acgcccgtacagcgcgtagcgacacagggagcgtgcgcggacacacagcgtgaag 5631

|||||  
Db 787 AAGCTTATCAACCGGATGGGAGAGACACGGAGTTCGCCAGCAACACAGCTTAAG 728  
Qy 5632 acgctgtcacaacacagacacagcctccctacatacaagctcgatccgtgtgtccac 5691  
Db 727 ACGATGCTTACAGGAGAAAGACTTTCCCTACATCAGAGCTGCATCTGTGTCTCTC 668  
Qy 5692 cggagagagacggtgtcagcagctgaggtgaggtgacacacagacatgacagaaagaca 5751  
Db 667 GGTAAAGACAGCGTCTCTACGCGCCAGTGGAGGTGCATTCAGACATGCGAAGAAAGACA 608  
Qy 5752 cggagcctgagccttgcacacagcagcagcaccacagatgctaagatgtacagatgtg 5811  
Db 607 CGGAGCTGGCGCTGTGCCACGACAGACACCAACATGCTAAGATGTACAGATGTG 548  
Qy 5812 cttaagagctctgttagagggccacagctgaacacaggttccctgaggtgagcagaggttt 5871  
Db 547 CTTCAGGCTCTGTAGGGCCCAACCGTGAACAGAGGTCTCTGGAAGTGGCCAGGTGTT 488  
Qy 5872 ttagcagagatcccggaagaccacagctctccgcatcacaacaaatgacgctctgc 5931  
Db 487 TTACGAGAGATCCCGAGAGACCCCAAGCTTCCGGCATCAACAATATGCGGCTCTGC 428  
Qy 5932 ttaagagactctgcaagaataatgtgagatgctgtgcgaaaaataagggccctgattgg 5991  
Db 427 TTCAAGAGACTTCTGCAGAAATGTAGAGATGCTGCGAATAATAGCCCTGATTGGG 368  
Qy 5992 ccgagacagaagagatlacaccgctgagctgagcgaactactgcccgtcgagagagct 6051  
Db 367 CCGSACAGAAAGAGATACACCGTAGCTGAGCGCAACTCTCCGCTGCGGAGAGGT 308  
Qy 6052 ctgcagccctgtgcttaaccacagcgtgcccagctgtagtgacacccaccccgccctc 6111  
Db 307 CTGCAGCCCCCTGTCTTACCCACGCGCTGCCCAATGTAGAGACCCACCCCGGCTC 248  
Qy 6112 aggaactccttgacaagcagaatctccgaagcagaacctgagccacaagagacaa 6171  
Db 247 AAGAACTCCTTGAACAGAGCAAGTTTCGAAAGGACAGACCTGTAGGCCACAAAGSACAA 188  
Qy 6172 agctgtacctagagagacacagacacccgggctcagctgtgtgtgtgtgtgtgtgtgt 6231  
Db 187 AACTTACTTACAGGAGAACACACACCCGGGCTCAGCTGTCTGCTCGAGGGAGAGCTG 128  
Qy 6232 cccctgtgcccacatgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 6291  
Db 127 CCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 68  
Qy 6292 ctgtgtcccatctgt 6351  
Db 67 CTGTGTCCCATCTGT 8  
Qy 6352 taagcaa 6358  
Db 7 TAAACAA 1  
RESULT 13  
LOCUS BF569035 709 bp mRNA linear EST 12-DEC-2000  
DEFINITION BF569035 NIH\_MGC\_42 Homo sapiens cDNA clone IMAGE:4300412 3',  
mRNA sequence.  
ACCESSION BF569035  
VERSION BF569035.1 GI:11642415  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 709)  
NIH-MGC http://mgi.nci.nih.gov/.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.

Email: c9ababs-remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L16M155 row: e column: 21  
High quality sequence start: 3  
High quality sequence stop: 707.

BASE COUNT	120 a	212 c	223 g	154 t
ORIGIN				

Query Match	10.7%	Score 683.4;	DB 10;	Length 709;
Best Local Similarity	99.6%	Pred: No. 3.8e-113;		
Matches 706; Conservative	0;	Mismatches 1;	Indels 2;	Gaps 2;

QY	5604	ggaagctgcccgaagaacaagaagttaagacgcgtctcagcaacgcgcacgctctccctca	5063
Db	709	ggagctgcccgaagcaacacacgaacgtaagacgcgtctcagcaacgcgcacgctctccctca	650
QY	5664	cataaagatcgaatccgtgtgtgccaaccgggaagagaagatgtcgaagccagtggaagct	5723
Db	649	catcaagctcgcattccgttctgtgccaccggagagacggtgctgaagccagtggaagct	590
QY	5724	ggccatcgagagacatgcaagaagaagacacggyagctgyccttltgccacgagcagagacc	5783
Db	589	ggccatcgagagacatgcaagaagaagacacggagctgagccttggccaccgagcagagacc	530
QY	5784	accagaatcgaatgtcttaagaatgtgttcttaagagctctgtgaagccccacgctgaaaca	5843
Db	539	accagaatcgaatgtcttaagaatgtgttcttaagagctctgtgaagccccacgctgaaaca	470
QY	5844	gggtccctctggaagtgtgcccagatgttctttaagaagatccccgaagaaccccaagctct	5903
Db	469	gggtccctctggaagtggtgcccagatgttctttaagaagatccccgaagaaccccaagctct	410
QY	5904	ccggaatcaacaacaatgtcggtctgtctcaagaatctctcaagaatgttgatgagtc	5963
Db	409	ccggaatcaccacaaatgttggtggtctgtctcaagaatctctcaagaatgttgatgagtc	350
QY	5964	gctgtcggaataataagcccttgatcttggtccggaacgaagaagtagtaaccaacgtgagctgga	60233
Db	349	gctgtcggaataataagcccttgatcttggtccggaacgaagaagtagtaaccaacgtgagctgga	290
QY	6024	ggccaactactgcccgcctbgcggaagctctgcagccctgtcttaeccagcgcctgccc	6083
Db	289	ggccaactactgcccgcctbgcggaagctctgcagccctgtcttaeccagcgcctgccc	230
QY	6084	gctgtgtgcaaccccccaccccggtctcaagaatctcttgaacaagaagagttctcgaa	6143
Db	229	gctgtgtgcaaccccccaccccggtctcaagaatctcttgaacaagaagagttctccgaa	170
QY	6144	ggcgaactctgagcccacaagaacccaagctgtacctagaggaacacagaccggagct	6203

Accession	Sequence	Position
Db	GGCAGACCTCTGAGCCCAAGAGNCCAAAGTGTACTAGAGGAAACAGCACCCGGGCT	110
Qy	gacgtgtctgtgct-gcgaaaggagctgtccctgtgtccacatggtctgttgggttgaca	676
Db	CAGCTGTGTGCTGCGCAGGGGAGTGTGCATCATGGTCCCACTGGGTGTGGGTGACCA	50
Qy	cacttacttgggtcttggtccctctgtcccctgtgtcccaatctgtgac	6311
Db	CACGTACTTGGGGCTGGGCCCTTG-CCCTGTGTCCCACTGTGTGTC	2

RESULT	14
BIB822534	
LOCUS	
DEFINITION	B1822534 745 bp mRNA linear EST 04-OCT-2001 603035242P1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5176259 5' , mRNA sequence.
ACCESSION	B1822534
VERSION	B1822534.1 GI:15934084
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
NIH-MGC http://mgc.ncl.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)	
Contact:	Robert Strausberg, Ph.D.

BASE COUNT	147 a	238 c	215 g	145 t
ORIGIN				
Query Match		10.6%	Score 677.4	DB 10; Length 745;
Best Local Similarity		98.0%	Pred. No. 4.6e-112;	
Matches 728; Conservative		0;	Mismatches 11;	Indels 4; Gaps 4;

QY 3250 cgaagctccaccccttcctcaagcgaagcccggaagcccaagctgacagcagctgttctgaactg 3309

Db 1 CAGAGCTCCACCTTCTCCAGCCAAAGCCCCGGA -CCCAAGGTACACAGCATGTTCGAAGTG 59

QY 3310 a-gtggacattccggagagcagcagcctctctcagctgagctctctctctgaagagcttgagcct 3368

Db 60 AGGTGAGACATTCCGGCAGCAGCATCTTCAGTGGGGCTCTCTCTCAGCGAGCTGGCACT 119

QY 3369 gagccctggaactgagctgagctgaaggggcatctctgttgcacaaagaagccatcagtgagctg 3428



```
Db 120 GGGCCCTGAAACCTGAGGCTTAAGGGCATTCCTGTTGACACAGAGGCCATCAGTGTCTGT 179
QY 3429 gcaacagctctatgttggccatgacacatgacccccgctacgagcccaactgtgaagc 3488
Db 180 GCACAGGCTCTATGTGGCCATGACACTGACCCCGGTACGCCGAGCCACTGTGAAGGC 239
QY 3489 tctgtggccagactgtactgtccactgtcttgatgtgcaaggatccttgcacagcct 3548
Db 240 TCGTGTGGCCGACCTGTACTGTCACCTGATGATGATGATGATGATGATGATGATGATGAT 299
QY 3549 gcatgactgtctgagggccagctgacgagctgacagctgacagctgacagctgacagctgac 3608
Db 300 GCATGACTTGTGATGAGGCCAGGCCAGGTACAGGCTGACAGCTGCTGATGATGATGATGAT 359
QY 3609 cacagaagcgagagggagacatgacgaggtacacatcaccatccttctgtgacatgacatgc 3668
Db 360 CACAGAGGCGGAGGAGGACATGCGGGTACATCAACACCTCTGTGTGCGCATGGCCATTGC 419
QY 3669 tctgtggccagctgtgacccctgtgctccggggccacatcctcccaaggagccttc 3728
Db 420 TGTGTGCCCCCTTGTGCCCCCTGTGCCCCGACATCTCCAGGGGCCAACAGGCTTC 479
QY 3729 tctgagagctgtgtgcccctctctgtgacgagcagcagctgtgtgtgtgtgtgtgtgtgt 3788
Db 480 TCGCGCAGGCTGTGCCCCCTCTCTGTGCTGATGATGATGATGATGATGATGATGATGATGAT 539
QY 3789 ggtgtgtgaaacacacacacacacacacacacacacacacacacacacacacacacacacac 3848
Db 540 GGTGTGTGAAACACACACACACACACACACACACACACACACACACACACACACACACACAC 599
QY 3849 ccaagctgggagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3908
Db 600 CCAAGCTGGGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 657
QY 3909 gaaacagccttctgacacacacacacacacacacacacacacacacacacacacacacacac 3968
Db 658 GAAACAGCCTTCTGACACACACACACACACACACACACACACACACACACACACACACACAC 717
QY 3969 gacgctagaggaagcattctgtg 3991
Db 718 CGCGGTAGAGGAGGACCATCTGGG 740

RESULT 15
Bg678867 857 bp mRNA linear EST 01-MAY-2001
LOCUS Bg678867
DEFINITION 602624/67F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4749689 5',
ACCESSION Bg678867
VERSION Bg678867
KEYWORDS mRNA sequence.
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 857)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaab-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL0603 row: e column: 18
High quality sequence start: 4
High quality sequence start: 818.
Location/Qualifiers
```

```
source 1. 857
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4749689"
/clone_1ib="NCI_CGAP_Skn4"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (rtI phage-resistant)"
/note="Organ: skin; Vector: pCMV-Sport6; Site: 1: NCI;
Site: 2: Sali; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 209 a 254 c 247 g 147 t
ORIGIN

Query Match 10.6%; Score 673.8; DB 10; Length 857;
Best Local Similarity 92.2%; Pred. No. 2.1e-111;
Matches 790; Conservative 0; Mismatches 52; Indels 15; Gaps 7;

QY 5359 tcacacgagcttgagaggtcttacaagagagatgttgcaagcgtctgtgagatc 5418
Db 1 TTACACCGGCTGGAGGATTTCTACACGAGAGATTGGCCGACGCTGTGAGATTATC 60
QY 5419 aaagactcttaccctgtgtggaacagctcgaagcttgacacaaaggcctacacagatc 5478
Db 61 AAAGACTCTTAACCTGTGACAACTTCAAGCTTACTCACAAGGCTTACTATCAATC 120
QY 5479 acgtatgtgaaacgtacttctgacacacacacacacacacacacacacacacacacacac 5538
Db 121 ACGTATGTGAACCGTACTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
QY 5539 cgaacatagtgcttgcacatctctgttctgacacacacacacacacacacacacacacacac 5598
Db 181 CGCACTATGAGGCTTCCACATCTCTGTGTGACGCGCTTCAACGCGGAGGCGGAGGCGCA 240
QY 5599 cagcggagagctgcccagacacacacacacacacacacacacacacacacacacacacacac 5658
Db 241 CACGGGAGGCTGCCGAGCAACACACACACACACACACACACACACACACACACACACACACAC 300
QY 5659 ccccaatcaagactgcacatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5718
Db 301 CCTTACATCAAGACTCCCATCCGATGTCACACGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
QY 5719 gaagtgcacatcgagacacacacacacacacacacacacacacacacacacacacacacac 5778
Db 361 GAGGTGCCATCGAGGACATGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAG 420
QY 5779 gaccacacagatgttaagatgtctacagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5838
Db 421 GACCCACACAGATGCTAGATGCTACAGATGCTGCTCAGGCTCTGTAGAGCCACACCGTG 480
QY 5839 aaccaggtccctgtgagagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5898
Db 481 AACCA-GGTCCCTCGGAGGTGGCCAGAGTGTGTTTACACAGAGACCCGAGAGACCCCAAG 539
QY 5899 ctcttcgagatcaacaatgtcggtctgtc-ttaagagacttctgcaagaaatgtga 5957
Db 540 CTCTTCGGGATCAACAATGCGCTGCTTGAAGGACTTGTGCAAGAAATGTGA 599
QY 5958 gcatgctctcgagaaataagagccctgcatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 6016
Db 600 GAGTGGCTCTCGGAAATAAGGCGCTGATGCGGCGGACACAGAGAGATACACCGGG 659
QY 6017 -agctgagcgcaactctgcgcctgtgagagctgtgacgcccctgtacc----- 6070
Db 660 AGCTTTGAAGGCACTACTGCGGCTCGGGGAGGCTCTGAGGCGCTTACCCCAAGC 719
QY 6071 -agcgctgtccacagctgtgacacacacacacacacacacacacacacacacacacacac 6127
Db 720 GCGCTGCCCAACTGATGAGGACCCACCCACAGGCTTCAGGAATCTCTTGGAACA 779
QY 6128 gagcaag--tttcgaaagcagacactgtgacccacaaga--ccaagctgtaccata 6184
||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
```

Db 780 GAGCAAGGTTCCGAGGGGAGACCTCTGAGCCACAGAGGCCAAAGCTGGACCTAGA 839  
QY 6185 gnaaccagcaccg9gc 6201  
| | | | |  
Db 840 AGGAACCAAGCACCGGCGC 856

Search completed: October 5, 2002, 04:27:31  
Job time: 12634 sec

Mon Oct 7 12:35:17 2002

OM of: US-09-736-968a-2 to: GenEmbl.\* out\_format : pfs

Date: Oct 5, 2002 3:01 PM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=frame-pzn.model -DEV=xlh  
-Q=gen2.1/uspto\_spool/us09736968/runat.02102002.145334.1333/app\_query.fasta.1.2139  
-DB=GenEmbl -QEXT=fastcap -SUFFIX=rgx -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOCP=0.000 -LOOEXT=0.000 -GAPOP=6.500  
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -GAPOP=6.000  
-DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=bloms62 -TRANS=human40.cdi  
-LIST=45 -DOCAIGN=200 -THR\_SCORE=pcr -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTPR=pts -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US09736968@cgn1.1.10376  
-NCPU=6 -ICPU=3 -LONGJOB -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-736-968a-2

Database: GenEmbl.\*

Database lengths: 1797656

Search time (sec): 8046.500000

score\_list:

Sequence	Strd Orig	ZScore	Escore	len	Documentation
gb_pat:AX173175	+ 10569.00	9403.15	0.0	6372	AX173175 Sequence 1 from Pat
gb_pat:AB037816	+ 8334.00	7414.12	0.0	4886	AB037816 Homo sapiens mRNA for
gb_pat:AX172874	+ 7198.00	6598.87	0.0	6828	AX172874 Sequence 1 from Pat
gb_pat:AX174559	+ 5258.50	4669.51	9.8e-252	7215	AX174559 Sequence 1 from Pat
gb_ov:AY016023	+ 4734.00	4170.51	6.0e-224	167215	AY016023 Spherooides nephe
gb_pat:AB051558	+ 4453.50	3957.31	4.6e-212	4311	AB051558 Homo sapiens mRNA for
gb_htg:AC017375	+ 4433.00	3910.59	1.8e-209	73515	AC017375 Drosophila melanog
gb_pat:AC008318	+ 4423.00	3893.48	1.6e-208	16626	AC008318 Drosophila melanog
gb_pat:AC003590	+ 4423.00	3887.52	3.5e-208	301639	AC003590 Drosophila melanog
gb_pat:AC008335	+ 4419.00	3530.80	1.4e-210	2825	BC008335 Homo sapiens, clone
gb_pat:AX172880	+ 4273.50	3797.16	3.8e-203	4200	AX172880 Sequence 107 from Pat
gb_pat:AX172960	+ 4273.00	3796.85	4.0e-203	4144	AX172960 Sequence 1 from Pat
gb_pat:AX172932	+ 4273.00	3796.85	3.7e-203	4144	AX172932 Sequence 59 from Pat
gb_pat:AK055905	+ 3626.50	3322.92	3.0e-171	3327	AK055905 Homo sapiens cDNA FL
gb_pat:AX172978	+ 3508.50	3118.00	2.6e-165	3253	AX172978 Sequence 105 from Pat
gb_pat:AX173233	+ 3326.00	2859.75	1.7e-156	2149	AX173233 Sequence 59 from Pat
gb_pat:AK024436	+ 3291.00	2920.75	1.5e-154	4577	AK024436 Homo sapiens mRNA for
gb_pat:AX174571	+ 2850.50	2529.49	1.5e-132	4026	AX174571 Sequence 3 from Pat
gb_pat:AX174664	+ 2850.50	2529.49	1.5e-132	4027	AX174664 Sequence 96 from Pat
gb_pat:AK055493	+ 2838.00	2523.33	3.4e-132	2452	AK055493 Homo sapiens cDNA FL
gb_ro:BC021414	+ 2821.00	2510.74	1.7e-131	1900	BC021414 Mus musculus, clone
gb_pat:AX173022	+ 2415.00	2136.67	1.2e-110	6454	AX173022 Sequence 1 from Pat
gb_pat:AX235048	+ 2360.00	2086.14	7.7e-108	7506	AX235048 Sequence 7 from Pat
gb_pat:AX174662	+ 2172.50	1932.25	2.9e-99	2014	AX174662 Sequence 94 from Pat
gb_pat:AC019102	+ 2054.50	1823.25	3.3e-93	2904	AC019102 Homo sapiens, Simla
gb_pat:AK055401	+ 2001.50	1777.27	1.2e-90	2607	AK055401 Homo sapiens cDNA FL
gb_pat:AC009005	+ 1996.50	1730.93	4.7e-88	205035	AC009005 Human chromosome 14
gb_pat:AB028981	+ 1966.50	1737.96	1.9e-88	5862	AB028981 Homo sapiens mRNA for
gb_pat:AC019102	+ 1726.50	1492.27	9.2e-75	139887	AC019102 Homo sapiens chromosome 14
gb_pat:CEBPA665	+ 1702.50	1483.72	8.8e-74	38886	CEBPA665 Human chromosome 14
gb_pat:AX173028	+ 1627.50	1438.76	8.8e-72	4391	AX173028 Sequence 7 from Pat
gb_pat:AX173118	+ 1609.50	1422.71	6.9e-71	4393	AX173118 Sequence 7 from Pat
gb_pat:AC008354	+ 1594.00	1372.89	4.1e-68	159350	AC008354 Drosophila melanog
gb_pat:AC003617	+ 1594.00	1368.09	7.3e-68	257224	AC003617 Drosophila melanog
gb_pat:AX058220	+ 1552.00	1332.45	7.3e-66	215105	AX058220 Sequence 90 from Pat
gb_pat:AX058220	+ 1541.50	1363.31	7.1e-67	3899	AX058220 Sequence 90 from Pat
gb_pat:AC011472	+ 1465.50	1296.35	1.0e-56	122321	AC011472 Homo sapiens chrom
gb_pat:AK056684	+ 1361.50	1168.35	3.9e-56	3152	AK056684 Homo sapiens cDNA FL
gb_pat:AX058192	+ 1197.00	1060.34	1.1e-50	2610	AX058192 Sequence 62 from Pat

gb_pat:AC009000	+ 1153.00	994.51	4.9e-47	37107	AC009000 Homo sapiens chro
gb_pat:AF465831	+ 1116.00	979.99	3.1e-46	5896	AF465831 Arabidopsis thalia
gb_pat:AK054649	+ 1091.50	967.60	1.5e-45	2299	AK054649 Homo sapiens cDNA
gb_pat:AC097166	+ 1089.50	927.87	2.5e-43	101121	AC097166 Rattus norvegicu
gb_pat:AF194407	+ 1080.00	963.52	2.6e-45	1242	AF194407 Homo sapiens unkno

seq\_name: gb\_pat:AX173175

seq\_documentation\_block:

LOCUS AX173175 6372 bp DNA linear PAT 03-JUL-2001

DEFINITION Sequence 1 from Patent WO0142295.

BASE COUNT  
ORIGIN

1273 a 2050 c 1849 g 1200 t

102aa protein B

## alignment\_scores:

Quality: 10569.00 Length: 2047  
Ratio: 5.163 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-736-968a-2 x AX173175 ..

Align seg 1/1 to: AX173175 from: 1 to: 6372

```
1 MetAlaIaSerGIuArGArGAlaPheAlaHisTylLeuAsnArgThrVa 17
  |||
13 ATGGCTGCTCTCCAGAGCCCGCCCTTCCGACAAAGATCAACAGCGGT 62
  |||
17 LAlaIaGIuValArgLysGIuValSerArgGIuArgSerGIySerProH 34
  |||
63 GCGCGAGAGGTGCGAAGCAGGTCTCCGGAACGACAGTGTCCCTCC 112
  |||
34 LsSerArgArgCysSerSerSerLeuGIyValProLeuThrGIuVal 50
  |||
113 ACTCCAGAGGCTGCGACAGCTCCCTGGGGTCCACTGACTGAAATT 162
  |||
51 ValGIuProLeuAspPheGIuAspValLeuLeuSerArgProProAspAl 67
  |||
163 GTGAGGCCCTGGACTTTGAGAGATGACTTCTGAGCCGCGCACAGATGC 212
  |||
67 AGIuProGIyProLeuArgAspLeuValGIuPheProAlaAspAspLeuG 84
  |||
213 TGAGCCCGGGCCCTCAGAGAGCTGGAGATTCCAGCTGATGACTTGG 262
  |||
84 LuLeuLeuLeuGIuProArgLyuCysArgThrThrGIuProGIyLePro 100
  |||
263 AGCTGTGCTGCAGCCCGGGAAATGCCGACACAGGAGCCCGGGATCCCC 312
  |||
101 LysAspGIuLysLeuAspAlaGIuValArgAlaValGIuMetYrL 117
  |||
313 AAGGATGAAAACTGATGCCAGGTGAGGCGCGCGTGAGATGTATAT 362
  |||
117 eGIuAspTrValIleValHisArgArgTYrGIuTYrLeuSerAlaAlAT 134
  |||
363 TGAGGACTGGGTCAATGTCCACAAAGATATCAGTACCTGATGCGAGCAT 412
  |||
134 YrSerProValThrThrAspThrGIuArgGIuArgGIuLysGIuLeuPro 150
  |||
413 ACAGCCCGTACACACAGACACACAGCGGAGACAGAGGCGCTCC 462
  |||
151 ArgGIuValPheGIuGIuAspAlaSerGIyAspGIuArgSerGIyProGI 167
  |||
463 CGCCAGGCTTTGAGCAGATGCTTGGAGAGAGAGTCCGGCCCTGA 512
  |||
167 uAspSerAspAspSerArgArgLysSerGIySerProGIuAspThrProAl 184
  |||
513 GGAAGTCGAATGACTCCGGCGGTGGCTCCGCGAAGACACCCCTC 562
  |||
184 rGSerSerGIyAlaSerSerIlePheAspLeuArgAsnLeuAlaAsp 200
  |||
563 GAACAGATGGTGGCTTAGCATCTTGACCTGAGGAACCTGGACGTGAC 612
  |||
201 SerLeuLeuProSerLeuLeuGIuArgAlaAlaProGIuAspValAspPr 217
  |||
613 TCATTGGCTGCTCTGTGTAGAGCGGGCGGCCCAAGATGTGGACCG 662
  |||
217 gArGAsnGIuThrLeuArgArgGIuHisArgProProAlaLeuThrL 234
  |||
663 GCGCAATGAACCTTCGAGCGGACGACCGGCCCGGCTGCTCAACC 712
  |||
234 euTYrProAlaProAspGIuAspGIuAlaValGIuArgCysSerArgPro 250
  |||
713 TCTATCCCGGCACTGACGAGAGATGAAGCCGTGAGACGCTGTACCC 762
  |||
251 GIuProProArgGIuHisPheGIyGIuArgIleLeuValLysCysLeuSe 267
  |||
763 GAGCACCAGCGGACGACTTTGGACAAGAGATTGTGTCAAGTGTGTGTC 812
```

```
267 rLeuLysPheGIuIleGIuIleGIuProIlePheGIyIleLeuAlaLeuT 284
  |||
813 GCTCAATTGCAAGATTGAATTAGCCCACTTTGGGATCTTGGCTCTGT 862
  |||
284 YrAspValArgGIuLysLysLysIleSerGIuAsnPheTYrPheAspLeu 300
  |||
863 ATGATGTGCGGGAGAAAAAGAAATCTCGAGAACTTCTACTCTGCACCTG 912
  |||
301 AsnSerAspSerMetLysGIyLeuLeuArgAlaHisGIyThrHisProAl 317
  |||
913 AACCGAGCTCCATGAAGGGGCTGCTTCCGGCTCATGCGCACCCCTCC 962
  |||
317 aIleSerThrLeuAlaArgSerAlaIlePheSerValThrTYrProSerP 334
  |||
963 CATCTCCACCTGCGCCGCTCTCCATCTTCTGTGTGACTTACCTCCAC 1012
  |||
334 roAspIlePheLeuValIleLysLeuGIuLysValLeuGIuArgGIuAsp 350
  |||
1013 CTGACATCTTCTGTGATCATCAAGTTGGAGAGGTCTTACGCAAGGGAC 1062
  |||
351 IleSerGIuCysCysGIuProTYrMetValLeuLysGIuValAspThrAl 367
  |||
1063 ATCAGTGAAGTGTGTGAGCTCTACATGCTGTGAAAGAGTGGACACAC 1112
  |||
367 aLYsAsnLysGIuLysLeuGIuLysLeuArgLeuAlaIleGIuArgPheC 384
  |||
1113 CAAGACAAAGAGAGAGCTAGAGAGAGTGCCTGGCGCGCGACAGTCTT 1162
  |||
384 YrThrArgLeuGIyArgTYrArgMetProPheAlaTrpThrAlaValHis 400
  |||
1163 GCACCCCGCTGGGGCGGTACCGCATGCCCTTCCCTGGAGCGCGCGTGCAC 1212
  |||
401 LeuAlaAsnIleValSerSerAlaGIuGIuLeuAspArgAspSerAsp 417
  |||
1213 TTGGCCAAACTGCTGACACAGCTGGGCACTGACCGGAGACTTGTACTC 1262
  |||
417 rGIuGIyGIuArgArgProAlaTrpThrAspArgArgArgArgGIyProG 434
  |||
1263 GGAGGGGAGAGCGCGCGGACGCTGGAGACGCCCGCCCTCGGGGGCCCC 1312
  |||
434 LAspArgAlaSerSerGIyAspAspAlaCysSerPheSerGIyPheArg 450
  |||
1313 AGGACCGGGCGAGTAGTGGGAGCAGCGCTTGCAGCTTCTGTGGCTTCCGT 1362
  |||
451 ProAlaThrLeuThrValThrAsnPhePheLysGIuAlaGIuArgLe 467
  |||
1363 CCAGCCACGCTTAAGTGCACAAACTTCTTAAACAGAGAGGCTGAGCGACT 1412
  |||
467 uSerAspGIuAspLeuPheLysPheLeuAlaAspMetArgArgProSerS 484
  |||
1413 CAGTGAAGAGGACTCTTCAAGTTCTGTGGTGAACATGAGAGGCCCTGCT 1462
  |||
484 eXLeuLeuArgArgLeuArgProValThrAlaGIuLeuLysIleAspIle 500
  |||
1463 CCTGTGTGCGGGCGACTACGTCTGTGACTGCCAGCTCAAGATCGACATT 1512
  |||
501 SerProAlaProGIuAsnProHisPheCysLeuSerProGIuLeuLeuH 517
  |||
1513 TCTCCGGCTCTGAAATCCCACTTGTGCTCTCCCTGAGAGCTGCTTAA 1562
  |||
517 sIleLysProTYrProAspProArgGIyArgProThrLysGIuIleLeuG 534
  |||
1563 TATCAACACCTTACCGGAGCCAGGGGCGGCGCCACCAAGAGATTTCTGG 1612
  |||
534 LuPheProAlaArgGIuValTYrAlaProHisThrSerTYrArgAsnLeu 550
  |||
1613 AGTTCGCCCGCGGAGAGTATGCCCCCATTCAGCTACAGGAACCTGG 1662
  |||
551 LeuTYrValTYrProHisSerLeuAsnPheSerSerArgGIuArgIleSerVa 567
  |||
1663 CTGTAGCTGTACCGGACAGCTCACTCACTTACAGAGCGCGCAGGGCTCGGT 1712
```

```

567  lArGAsnLeuAlaValArgValGlnTyrMetThrGlyGluAspProSerG 584
    |||
1713  GCGCAACCTTGGCTGTGGAGTGCATGATGACAGCGAGAGACCCAGCC 1762
    |||
584  lAlaLeuProValIlePheGlyLysSerSerCysSerGluPheThrArg 600
    |||
1763  AGGCTCTCGCGTCACTTTTGGCAAGTCCAGCTCAGTAATTACCCGC 1812
    |||
601  GluAlaPheThrProValValTyrHisAsnLysSerProGluPheTyrG 617
    |||
1813  GAGGCTTCACACCGGTGGTCTACCATTAACAAGTCCCGGAGTTCTAGA 1862
    |||
617  uGluPheLysLeuHisLeuProAlaCysValThrGluAsnHisHisLeu 634
    |||
1863  GGAGTTCAGCTGCATCTTCCAGCTGGGTGACAGAGAACCATCACTGC 1912
    |||
634  eupheThrPheTyrHisValSerCysGlnProArgProGlyThrAlaLeu 650
    |||
1913  TGTTCACCTTCTACCATGTCAGTGCCAGCCCGCGGCGGACCTGCCCTG 1962
    |||
651  GluThrProValGlyPheThrTyrIleProLeuLeuGlnHisGlyArgLe 667
    |||
1963  GAGACACCCCGGGGCTTACTTGGATCCCACTGCTGCAGCACGGGCGCT 2012
    |||
667  uArgThrGlyProPheCysLeuProValSerValAspGlnProProSer 684
    |||
2013  GAGAGACGGCCCCCTTCTGTCTCCAGTGTGTGGACACGCGCGGCCCA 2062
    |||
684  eTyrSerValLeuThrProAspValAlaLeuProGlyMetArgTyrVal 700
    |||
2063  GCTATTCGCTGCTCACACCCGATGTGGCGGTCCGGGCGATGGCGTGGTG 2112
    |||
701  AspGlyHisLysGlyValPheSerValGluLeuThrAlaValSerSerVa 717
    |||
2113  GACGGTCAACAAGGCGGTTCAGTGTGAGCTCACACCGGTCTCTGTGT 2162
    |||
717  lHisProGlnAspProTyrLeuAspLysPhePheThrLeuValHisVal 734
    |||
2163  GCACCCCGAGACCCCTACCTGCGCAAAATCTTACCCTGGTGGACGTCC 2212
    |||
734  euGluGluGlyAlaPheProPheArgLeuLysAspThrValLeuSerGlu 750
    |||
2213  TGGAGAGAGGAGGCTTCCCATTTCCGGCTCAAGGACACTGTGTGAGGAG 2262
    |||
751  GlyAsnValGluGlnGluLeuArgAlaSerLeuAlaAlaLeuArgLeuAl 767
    |||
2263  GGCACAGTGGAGCAGGAGCTGGCGGCCAGTCTTGACACTGGCGCTGGC 2312
    |||
767  aSerProGluProLeuValAlaPheSerHisHisValLeuAspLysLeuV 784
    |||
2313  CAGCCCCGAAACCCCTTGTGGCTTCTCCACACAGTGTGAGCAAGCTCG 2362
    |||
784  aLArgLeuValIleArgProProIleIleSerGlyGlnIleValAsnLeu 800
    |||
2363  TGGCTGTGGTCATCAGGCCCGCGCATCATAGTGGCCGATTTGGAACCTG 2412
    |||
801  GlyArgGlyAlaPheGluAlaMetAlaHisValValSerLeuValHisAr 817
    |||
2413  GGGCGTGGAGGCTTGAAGCAATGGCCCATGTAGTCAAGCTTGTTCACCG 2462
    |||
817  gSerLeuGluAlaAlaGlnAspAlaArgGlyHisCysProGlnLeuAla 834
    |||
2463  GAGCTCGAGGCAACCCAGATGCCCGCGGTCACTGGCCACAGCTGGCTG 2512
    |||
834  lArgTyrValHisTyrAlaPheArgLeuProGlyThrGluProSerLeuPro 850
    |||
2513  CCTACGTCCACTAGCTTTCGCTTCTGCACTGAGCCAGCCAGCTCCCG 2562
    |||
851  AspGlyAlaProProValThrValGlnAlaAlaThrLeuAlaArgGlySe 867
    |||
2563  GATGGGGCCCCCTCAGTACAGTGCAGGCTGCCACACTGGCCGCTGGCTC 2612
    |||
867  rGlyArgProAlaSerLeuTyrLeuAlaArgSerLysSerLysSerSers 884
    |||
2613  TGGTGGCCCCGAGGCTCTACTGGCGGTTCACAGACATCAGACAGA 2662
    |||
884  eArgAsnProAspLeuAlaValAlaProGlySerValAspAspGluValSer 900
    |||
2663  GCAACCTGACCTCGGCCGCGTGGCCCTGGGCTGTGTGATACAGAGTTTCC 2712
    |||
901  ArgIleLeuAlaSerLysLeuLeuHisGluGlyLeuAlaLeuGlnTyrVa 917
    |||
2713  CGCATCTGGCCAGCAAGCTGCTTCAGAGAGAGTGGCTCTCAGAGGGGT 2762
    |||
917  lValSerSerSerAlaValArgGluAlaIleLeuGlnHisAlaThrPheP 934
    |||
2763  GGTACAGCAGAGTGGCCGTACGGGAGGSCATCTCCAGACAGCTGGTGT 2812
    |||
934  hePheGlnLeuMetValLysSerMetAlaLeuHisLeuLeuGlnGly 950
    |||
2813  TCTTCAGGTCATGATGAGAGATGCGCTGCACCTGCTGTGGCCAG 2862
    |||
951  ArgLeuAspThrProArgLysLeuArgPheProGlyArgPheLeuAspAs 967
    |||
2863  CGACTAGACACACCCCGCAAGCTGCGCTTCCCGGACGCTTCTGAGAGA 2912
    |||
967  pIleThrAlaLeuValGlySerValGlyLeuGlnValIleThrArgValH 984
    |||
2913  CATCACTGCTTGTGTGGTCTGTGGCTGGAGGTATCACTACCCGTTGTC 2962
    |||
984  lLysAspValGluLeuAlaGlnHisLeuAsnAlaSerLeuAlaPhePhe 1000
    |||
2963  ACAAGATGTGAGCTGGCGCGAGACACTCAAGCAGCCTGGCTTTCTTC 3012
    |||
1001  LeuSerAspLeuLeuSerLeuValAspArgGlyPheValPheSerLeuVa 1017
    |||
3013  CTCAGTGAACCTTCTGTCCCTGTGTGACCGGGGCTTTGTCTCAGCTGGT 3062
    |||
1017  lArgAlaHisTyrLysGlnValAlaThrArgLeuGlnSerSerProAsp 1034
    |||
3063  CCGGGGCCACTACAAAGCAGGTGGCCAGCGGCTCCAGTGTCCCTTAATC 3112
    |||
1034  roAlaAlaLeuLeuThrLeuArgMetGluPheThrArgIleLeuCysSer 1050
    |||
3113  CAGCACCCCTGGTGAACCCGCGCATGTGATTCACCCGCACTCTGTGGACG 3162
    |||
1051  HisGlnHisTyrValThrLeuAsnLeuProCysCysProLeuSerProPr 1067
    |||
3163  CAGGACACTAGTGTGACCTCAACCTCCCTGCTGCCCCCTGTCACTCC 3212
    |||
1067  oAlaSerProSerProSerValSerSerThrThrSerGlnSerSerThnP 1084
    |||
3213  AGCTTGGCCCTCCCTCTGTGTCTCCACACACTCCAGAGCTTCACCT 3262
    |||
1084  heSerSerGlnAlaProAspProLysValThrSerMetPheGluLeuSer 1100
    |||
3263  TCTCCAGCCAAAGCCCGGAGACCCCAAGGTGACAGCATGTTCACACTGAGT 3312
    |||
1101  GlyProPheArgGlnGlnHisPheLeuAlaGlyLeuLeuThrGluLe 1117
    |||
3313  GGACCATTCGGGACAGACATTCCTAGTGGGTCTCGTGCAGGAGACT 3362
    |||
1117  uAlaLeuAlaLeuGluProGluAlaGluGlyAlaPheLeuLeuHisLysL 1134
    |||
3363  GGCACGTGGCCCTCGAACCCTGAGGCTGAAGGGCATTTCTGTGGCAAGA 3412
    |||
1134  yAlaAlaIleSerAlaValHisSerLeuLeuCysGlyHisAspThrAspPr 1150
    |||
3413  AGGCCATGATGCTGTGCACACAGCTGATGTGGGCATAGACACTACCCG 3462
    |||
1151  ArgTyrAlaGluAlaThrValLysAlaArgValAlaGluLeuTyrLeuP 1167
    |||
3463  CGCTAGCGCGGAGGCACTGTGAAGCTGTGTGGCGGAGCTGTACTCTCC 3512
    |||
1167  oLeuLeuSerIleAlaArgAspThrLeuProArgLeuHisAspPheAlaG 1184
    |||

```

3513 ACTGCTTCGATGTCACGGATACCTTTCACAGCGCTGCATGACTTTCGTG 3562  
1184 IUGLYProGlyIArgSerArgLeuAlaSerMetLeuAspSerAspThr 1200  
3563 AGGGCCCGAGGTCAGGGTCAGAGCTGGCCCTCAAGCTTGACTGCAGACACA 3612  
1201 GUGUGlyGUGlyAspIleAlaGlyThrIleAsnProSerValAlaMetAl 1217  
3613 GAAGCGCAAGGGGACATTCGGGTACCATCAACCCCTCTGCGCCATGGC 3662  
1217 aIleAlaGlyGlyProLeuAlaProGlySerArgAlaSerIleSerGlnG 1234  
3663 CATTCGCTGGTGGCCCTTCAGCCCTCGGCTCCGGGCGACATCTCCAGG 3712  
1234 LYPProThrAlaSerArgAlaGlyCysAlaLeuSerAlaGlnSerSer 1250  
3713 GGCACCAAGCGCTTCTCGCGAGCGTGGCTCTCTGCTGCACTCAAGC 3762  
1251 ArgThrLeuLeuAlaCysValLeuTrpValLeuLysAsnThrGluProAl 1267  
3763 CGGACCTTGTGGCGGTGTGTGCTGTGGGTGTGAAAAACACGAGCGGCG 3812  
1267 aLeuLeuGlnArgTrpAlaThrAspLeuThrLeuProGlnLeuGlyArgL 1284  
3813 GCTCTCGCAGCGCTGGGCGACTGACCTGACCTCCCGACGCTGGAGCTC 3862  
1284 euleuAspLeuLeuThrLeuCysLeuAlaAlaPheGluTrpLysGlyLys 1300  
3863 TGTTCGACTTGTGCTGACTTTCCTTGGCTGGCTTTCAGTACAAAGGGAAA 3912  
1301 LysAlaPheGluArgIleAsnSerLeuThrPheLysLysSerLeuAspMe 1317  
3913 AAGGCTTTCGAACGCATCAACAGCCTCACATTCAAAAAATCTCGATAT 3962  
1317 tLysAlaArgLeuGlnGluAlaIleLeuGlyThrIleGlyAlaArgGln 1334  
3963 GAAGGCGGCGCTAGAGAAAGCCATTCGGGTACATCGGAGCTCGACAG 4012  
1334 IuMetValArgArgSerArgGluArgSerProPheGlyAsnProGluAsn 1350  
4013 AATGCTTCGGGGAAGCTGTGAGAGAGACCCGTTTGGGAATCCAGAGAAC 4062  
1351 ValArgTrpAlaGlySerValThrHisTrpLysGlnThrSerAspArgVa 1367  
4063 GTGCGCTGGCGAAGAGCGTCACACACTGGAAGAAACCTCAGACCGCGT 4112  
1367 AspLysThrLysAspGluMetGluHisGluAlaLeuValGluGlyAsnL 1384  
4113 GGACAAAGCCAAAGGATGAATGGAACACGAGGCTTGGTGGAAAGGAAAC 4162  
1384 euAlaThrGluAlaSerLeuValIleLeuAspThrLeuGluIleIleVal 1400  
4163 TGGCAACCGAGGCAAGCCTAGTGGTTTGGACACACTGGAGATCATCGTG 4212  
1401 GlnThrValMetLeuSerGluAlaArgGlnSerValLeuGlyAlaValLe 1417  
4213 CAGAGGCGTATGCTTTCAGAAAGCCGGGAGAGCGCTTGGGGGCAAGTCT 4262  
1417 uLysValIleLeuTrpSerLeuGlySerAlaGlnSerAlaLeuPheLeu 1434  
4263 GAAGCTTGTGCTGTACAGCTCGGGCAGTGGCCAGAGTCCCTCTTCTTGC 4312  
1434 InHisGlyLeuAlaThrGlnArgAlaLeuValSerLysPheProGluLeu 1450  
4313 AGCATGGCGCTGGCCACCAGAGGGCGCTTGTCCAAAGTCCCGAGGCTG 4362  
1451 LeuPheGlnGluAspThrGluLeuCysAlaAspLeuCysLeuArgLeuLe 1467  
4363 CTGTTTCGAGGAGACAGAGCTGTGTGCGACCTGTGCTGAGAGCTCTCT 4412  
1467 uArgHisCysGlySerArgIleSerThrIleArgThrHisAlaSerAlas 1484  
4413 ACGACACTGTGTGACGCCCATTCACACCATCCGACGCGACCGCGGCT 4462  
1484 eLeuThrLeuLeuMetArgGlnAsnPheGluIleGlyHisAsnPheAla 1500  
4463 CGCTGTACTGCTCATGTCGACAGAACTTCGAGATCGGCCACAACCTTTCGC 4512  
1501 ArgValLysMetGlnValThrMetSerLeuSerSerLeuValGlyThrTh 1517  
4513 CGTGTAGATGACAGGTTCACATGTCTCTCTGCTCCCTGGTGGGAGCGAC 4562  
1517 rGlnAsnPheSerGluGlnHisLeuArgArgSerLeuLysThrIleLeuT 1534  
4563 GCAGAACTTCAGGAAGAGACACCTGGAGGTTGACTTCACTCAAAACATCTCA 4612  
1534 hTrpAlaGlnGluAspMetGlyLeuArgAspSerThrPheAlaGlnGln 1550  
4613 CCTATGCTAGAGAGACATGGGGCTCGGGACAGACACTTCACAGACGAG 4662  
1551 ValGlnAspLeuMetPheAsnLeuHisMetIleLeuThrAspThrValLy 1567  
4663 GTCAGAGACCTGATGTTCAACCTGCACATGATCTGACGACGACGAGTGA 4712  
1567 sMetLysGlnHisGlnGluAspProGluMetLeuIleAspLeuMetTrpA 1584  
4713 GATGAAGGAACACAGAGAGACCTGAGATGCTCATCGACCTCATGTACA 4762  
1584 rGlnAlaArgGlyTrpGlnGlySerProAspLeuArgLeuThrTrpLeu 1600  
4763 GAATTCGCCGGGCTACCAAGGCTCACCGAGCTTCGCTGACCGTGGTGG 4812  
1601 GlnAsnMetAlaGlyLysHisAlaGluLeuGlyAsnHisAlaGluAlaL 1617  
4813 CAGAACATGGCCGGGAGACCGGGAGCTGGGGCAACACAGCGGAGCGCC 4862  
1617 aGlnCysMetValHisAlaAlaAlaLeuValAlaGluTrpLeuAlaLeuL 1634  
4863 CCAGTGCATGTGTGCACGGCGCCCTCGTGGCTGAGTACCTCGCCCTGC 4912  
1634 euGluAspGlnArgHisLeuProValGlyCysValSerPheGlnAsnIle 1650  
4913 TCGAGGACACAGCCGACCTGCGCTGGCTGGCTTCCTCCAGAAACATC 4962  
1651 SerSerAsnValLeuGlnGlnSerAlaIleSerAspAspIleLeuSerPr 1667  
4963 TCATCCACAGCTGTAGAGAGATCCGCATCTCCGAGAAATCTGTTCGCC 5012  
1667 oAspGlnGlnGlyPheCysSerGlyLysHisPheThrGluLeuGlyLeuV 1684  
5013 CGACGAGAGAGGCTTCTGCTCCGGGAAGCATTCATGAGTGGGGCTGG 5062  
1684 aGlyLeuLeuGlnGlnAlaAlaGlyTrpPheThrMetGlyGlyLeuTrp 1700  
5063 TAGGTTTGTGGAAAGAGCGCGGCTACTTCAACCTGGCGGGGCTTAC 5112  
1701 GluAlaValAsnGluValTrpLysAsnLeuIleProIleLeuGlnAlaH 1717  
5113 GAGGCGGTGATGATGAGTCTTACAGAACTCATCCCATCTCGAAAGCCA 5162  
1717 sArgAspTrpLysLysLeuAlaAlaValaHisGlyLysLeuGlnAlaLAP 1734  
5163 CCGTGACTACAAAGAACCTGGCCGGGTGCACGGCAAACTGACGAGGGCT 5212  
1734 heThrLysIleMetHisGlnSerSerGlyTrpGluArgValaPheGlyThr 1750  
5213 TCACCAAGATCATGACACAGAGTTCGGGTGGAGGCGCTGTTCCGGAGG 5262  
1751 TrpPheArgValaGlyPheTrpGlyAlaHisAspPheLysAspLeuAspGlu 1767  
5263 TATTTCCGGGTGGCTTCTACGGCGGCCCACTTCGGTGAAGTGAAGACA 5312  
1767 ngLupheValTrpLysGluProSerIleThrLysLeuAlaGluIleSerH 1784  
5313 GGAAGTTTGTGTACAGAGCGCATCGATCACGAAGCTGGAGAGATCTCAC 5362

1784 tsargleugluGluuphetylthrgluargPheGluAspaspValValGlu 1800  
 |||||||  
 5363 ACCGGCTGGAGGGATTCTACAGGAGATTGGGACGACGCTGTTGAG 5412  
 |||||||  
 1801 llelleLysAspserTyrProValAspLysSerLysLeuAspSerGlnly 1817  
 |||||||  
 5413 ATATATCAAGACCTTACCTGTGGCAAGTCCAGCTTACCTGCACAAA 5462  
 |||||||  
 1817 salatYrllleGlnlleThrTyrValGluProTyrPheAspThrTyrGluL 1834  
 |||||||  
 5463 GGCTCTACATCCAGATCAGCTATGTGGAACCTGACTTGTATCTACAGAC 5512  
 |||||||  
 1834 euLysAspArgValThrTyrPheAspArgAspTyrGlyLeuArgThrPhe 1850  
 |||||||  
 5513 TCAGAGACCGGGGTGACCTTGTGACCGCACTATGGCTTGCACATTC 5562  
 |||||||  
 1851 LeupheCysThrProPheThrProAspGlyArgAlaHisGlyLeuPyr 1867  
 |||||||  
 5563 CTTGTTCTGCACCCGCTACCGCGGATGGCGCGCACACGGGAGACTGCC 5612  
 |||||||  
 1867 oGluGlnHisLysArgLysThrLeuLeuSerThrAspHisAlaPheProT 1884  
 |||||||  
 5613 CGAGCAACACCAAGCGTAAAGACGCTGCTCAGCACACGACGCTTCCCT 5662  
 |||||||  
 1884 YrllleLysThrArgLysValCysHisArgGluGluThrValleuThr 1900  
 |||||||  
 5663 ACATCAAGCTCGCATCCGCTGTGTGCCACCGGAGAGAGAGGCTGTGACG 5712  
 |||||||  
 1901 ProValGluValAlaIleGluAspMetGlnLysLysThrArgGluLeuAl 1917  
 |||||||  
 5713 CCAGTGGAGGTGGCCATCGAGGACATGCAGAAAGACACGGGAGACTGCG 5762  
 |||||||  
 1917 apheLarThrGluLysInAspProProAspAlaLysMetLeuGlnMetValL 1934  
 |||||||  
 5763 CTTTGCCACCGAGCAGAGCCACGATGCTAAGATGCTACAGATGAGTGC 5812  
 |||||||  
 5813 TTCAGAGGCTCTGTAGGGCCCAAGGCAAGGCTCCCTGGAGGTGGCC 5862  
 |||||||  
 1934 euGlnGlySerValGlyProThrValAsnGlnGlyProLeuGluValAla 1950  
 |||||||  
 1951 GlnValPheLeuAlaGluIleProGluAspProLysLysLeuPheArgHisH 1967  
 |||||||  
 5863 CAGGTGTTTTTAGTAGAGATCCCGGAAAGACCCCAAGCTTCCCGGCATCA 5912  
 |||||||  
 1967 sAsnLysLeuArgLeuCysPheLysAspPheCysLysLysCysGluAspA 1984  
 |||||||  
 5913 CAACAATATGGGGCTCTGCTTCAAGACTTCTGCAGAAATGTGAGATG 5962  
 |||||||  
 1984 laLeuArgLysAsnLysAlaLeuIleGlyProAspGlnLysGluTyrHis 2000  
 |||||||  
 5963 CGCTCGGAAAAATGAAAGCCCTGATGGCGGACACAGAGAGACTACAC 6012  
 |||||||  
 2001 ArgGluGluGluArgAsnTyrCysArgLeuArgGluAlaLeuGlnProLe 2017  
 |||||||  
 6013 CGTGAGCTGGAGCGCACTACTGCGGCTGCGGAGGCTGTGACGCCCT 6062  
 |||||||  
 2017 uLeuThrGlnArgLeuProGlnLeuMetAlaProThrProProGlyLeuA 2034  
 |||||||  
 6063 GCTTACCCAGGCGCTGCCCGAGCTGATGGACCCACCCCGCGGCTCA 6112  
 |||||||  
 2034 rGAsnSerLeuAsnArgAlaSerPheArgLysAlaAspLeu 2047  
 |||||||  
 6113 GGAACCTCCTTGAACAGACAAAGTTCCGAAAGGACAGACTC 6153  
 |||||||

seq\_name: gb\_AB037816

seq\_documentation block:

LOCUS AB037816 4886 bp mRNA linear PRI 14-MAR-2000  
 DEFINITION Homo sapiens mRNA for KIAA1395 protein, partial cds.  
 ACCESSION AB037816  
 VERSION AB037816.1 GI:7243170  
 KEYWORDS Homo sapiens brain cDNA to mRNA, clone\_lib:pbLuescriptII SK plus  
 SOURCE

ORGANISM Homo sapiens  
 clone:hj07927.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (sites)  
 AUTHORS Nagase,T., Kikuno,R., Ishikawa,K.I., Hiroseawa,M. and Ohara,O.  
 TITLE Prediction of the coding sequences of unidentified human genes.  
 XVI. The complete sequences of 150 new cDNA clones from brain which  
 code for large proteins in vitro  
 JOURNAL DNA Res. 7 (1), 65-73 (2000)  
 MEDLINE 20181126  
 REFERENCE 2 (bases 1 to 4886)  
 AUTHORS Ohara,O., Nagase,T. and Kikuno,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-JAN-2000) Osamu Ohara, Kazusa DNA Research Institute,  
 Laboratory of DNA Technology, 1532-3 Yata, Kisarazu, Chiba  
 292-0812, Japan (E-mail:cdnaInfo@kazusa.or.jp,  
 URL:http://www.kazusa.or.jp/huge/, Tel:+81-438-52-3913,  
 Fax:+81-438-52-3914)  
 FEATURES  
 source Location/Qualifiers  
 1..4886  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="hj07927"  
 /tissue="type="brain"  
 /clone\_lib="pbLuescriptII SK plus"  
 1..4884  
 /gene="KIAA1395"  
 <1..>4884  
 /gene="KIAA1395"  
 /note="Start codon is not identified."  
 /codon\_start=1  
 /product="KIAA1395 protein"  
 /protein\_id="BA092633.1"  
 /db\_xref="GI:7243171"  
 /translation="DARTMASERAFRAHKTINRYAAYKOVRESGSHSRCS  
 SLSGPLEVEYERPDDEVLISRPDAEGFLRDLVERPADLLELQPRECRTEEG  
 IPKDKLQYRAAYEMTIEDWVIVHRYOYLSAAYSPVTDLQREKQGLPRQVFO  
 DASGERSGPDSDNSRSGSPEDTPRSGASSIFDLRNLAOSLPLSLERAPEO  
 VDRNETLRQHRREPALLTLYPADDEAVRCSRPEPRPFQORILIVCOLSLKFEI  
 EIEPFGILAYDVERKKISENFYFDLSDSMKGLRAHGTHTAISTLASATFSYT  
 YSPDFIVITLKVLEVLQGGDISSECCPEKLEVDTKNKKLEKLELAEROTRG  
 RYRPFPAWTAHIANIVYSAGOLDRODSSEBERPANTDRRRRPPQASGDDACSF  
 SGRPATITVYNFKQAEKRLSDELEFKFMDMRPSLRLRPVPAQLKIDISPA  
 ENPHCSPELHLIKPIYDPGRPTKLETPAEVVAAPHTSYNNLYVYPHSINSS  
 RGSVRNLAIVQWMTGEDPSQALPVIFGKSCSEFTEAFTPVYHNKSPFEYEEK  
 LHLPACTVENHLLFTTFYHYVSCQPRGTALETGVGFIMPIPLQHRRLRTGFCIPVY  
 DQPPSYVLTPDVALPGMRWVDGKGVSELTAVASVHPDQYLLKFTVIVYLEE  
 GAFPRLDIVLSGYNVEOLRSLALRLASPRPLVAFSHNVDKTYRIVRPTIS  
 GQIVNLGARGAEAAHYSLYRSLLEAODRGCPOLAAVYHAFPLPTEBSLPDG  
 APPYVQANATLARGSGRPASLYLARSKSISSNPDLAVAPQSVDEVSRILASLHE  
 ELALQVWVSSAVREAILIQHMFEPOLNVSMALHLLGQRLDTPKLRPGRFLDI  
 TALVSGVLEIVTIVKHVLELAEHLNLSAFLFSLDSLVDRGVFLVRAHRYOVI  
 RLOSSPNPALITLMEFTRLTCSHEHYVTLNLCPLSPASPSVSTTSSQSF  
 SSOAPDPKVTSMFELSGPFRQOHLFAGLITELALALEPAREGAFILHKAASVHT  
 LCGHDPREVAETAKARVAELIPLUSTIARDTIPRIHDPABEGGQSRILASIMDSOT  
 EGBEDIACTINPSTAMALAGSPLAGSRASISQPTASRAGCLASNESSRTLACYL  
 VWLNTEPALLQKRWATDLTPQLGRLLDLYLCLAAEYGAKNFERINSLTPKSLD  
 MKARLEAIICTIGARQEMRERSRSPFGNENRWRKSVTHKOTSDSVAKTDEM  
 EHEALVGENLATEASLVVDLTLEIIVQTMLEARESVLGAIVLVYSLGSOALF  
 LOHGIATORALVSKPELLEPDEDELADCLRLRRLRGRISITRTHASASLYILMR  
 ONFELGNFAVRKQVQMSLSLVGTQNSSEHRLRSKLTITLYAEEDGILDSIFA  
 EYQVDLMFNHLITDVTYKKEHDEDEMLIDMTYRIARGIQGSPDLRLTWLQNMAGK  
 HAEIGNHAEACQVNAHA"

BASE COUNT 941 a 1605 c 1425 g 915 t  
 ORIGIN

alignment\_scores: Quality: 8334.00 Length: 1624  
 Ratio: 5.135 Gaps: 0  
 Percent Similarity: 99.938 Percent Identity: 99.938



alignment\_block:  
 US-09-736-968A-2 x AB037816 ..

Align seg 1/1 to: AB037816 from: 1 to: 4886

```

1  MetaIaIaSerGIuArGArGAlaPheAlaHIsLysIleAsnArGThrVa 17
   |||
13  ATGGCTGCTCCGAGCGCGCGCTTCGCGCACAAAGATCAACAGAGCGGT 62
   |||
17  IAlaIaGIuValArgLysGIuValSerArGIuArgSerGIySerProH 34
   |||
63  GCGCGAGAGGTGCGGAAGCAGGTGTCCGGGAACGACGTGCTCCCCC 112
   |||
34  IsSerSerArGIyCySerSerSerLeuGIyValProLeuThrGIuVal 50
   |||
113  ACTCCACAGCGCGCTGCAGCAGCTCCCTGGGGGTCCCACTGACTGAAGT 162
   |||
51  ValGIuProLeuAspPheGIuAspValLeuLeuSerArGIProProAsp 67
   |||
163  GTCGAGCCCTGGACTTTGAGGATGTACTCTGAGCGCGCACAGATGC 212
   |||
67  aGIuProGIyProLeuArGIAspLeuValGIuPheProAlaAspAspLeu 84
   |||
213  TGAGCCCGGGGCCCTCAGGAGCCTGTGAATTCACAGCTGATGACTTGG 262
   |||
84  IuLeuLeuLeuGIuProArGIuGIuCySerArGIThrGIuProGIyIlePro 100
   |||
263  AGCTGCTGCTGCAGCCCCGGGAATGCCGGACACAGCGAGCGCGGATCCCC 312
   |||
101  LysAspGIuLysLeuAspAlaGIuValArgAlaIaValGIuMetYrI 117
   |||
313  AAGGATGAAAAAATGAGATGCCACAGTAGAGGCGCGGTGAGATGTATAT 362
   |||
117  eGIuAspTrpValIleValHIsArGIyTrGIuThrLeuSerAlaIaT 134
   |||
363  TGAGGACTGGGTCAATTGTCCACAGAGGTATCACTACTGAGTGCACAT 412
   |||
134  yrSerProValThrThrAspThrGIuArGIuArgGIuLysGIyLeuPro 150
   |||
413  ACAGCCCGCTGCACACAGACACAGCGGGAGCGACAGAAGGGCTCCCC 462
   |||
151  ArgGIuValPheGIuGIuAspAlaSerGIyAspGIuArgSerGIyProGI 167
   |||
463  CGCCAGGCTTTGAGCAGGATGCTTGTGAGACGAGAGGTCCGGCCCTGA 512
   |||
167  uAspSerAsnAspSerArGIySerGIySerProGIuAspThrProA 184
   |||
513  GGACTCGAATGACTCCCGGCTGGCTGGGCTCCCGGAAAGACACCCCTC 562
   |||
184  rGIySerSerGIyAlaSerSerIlePheAspLeuArGIAsnLeuAlaIaAsp 200
   |||
563  GAAGCAGGTGGCTCTAGCATCTTCGACCTGAGAAACCTGGCAGCTGAC 612
   |||
201  SerLeuLeuProSerLeuLeuGIuArgAlaIaProGIuAspValaPar 217
   |||
613  TCATTGCTGCGCTCTGTGCTAGAGCGGGGCCCCCAAGATGTGGACGG 662
   |||
217  gArGAsnGIuThrLeuArGIuArgGIuHIsArGIProProAlaLeuLeuThr 234
   |||
663  GCGCAATAAACCTTCGACGCGACACGCGGCCCGGCTGCTCACCC 712
   |||
234  euThrProAlaProAspGIuAspGIuAlaValaGIuArgCySerSerArGIPro 250
   |||
713  TCTACCCCGGACCTGACGAGATGAAGCCGTGAAACGCTGAGCGGCCCA 762
   |||
251  GIuProProArGIuHIsPheGIyGIuArgIleLeuValLysLysLeu 267
   |||
763  GAGCCACCCCGGACCTTGGACAAAGGATCTTGGCAAGTGTCTGTCTC 812
   |||
267  rLeuLysPheGIuIleGIuIleGIuProIlePheGIyIleLeuAlaLeuT 284
   |||
813  GCTCAAGTTCAGAGATTGAATTTGAGCCCATCTTGGGATCTTGGCTGTCT 862

```

```

284  yrAspValArgGIuLysLysLysIleSerGIuAsnPheTrpPheAspLeu 300
   |||
863  ATGATGTGGGGGAGAAAAAGAGATCTCGAGAACTTCTACTGACCTG 912
   |||
301  AsnSerAspSerMetLysGIyLeuLeuArGIuHIsGIyThrHIsProAl 317
   |||
913  AACTCGAAGCTCATGAAGGGGCTGCTGGGTCTATGACCCACCCCTGC 962
   |||
317  aIleSerThrLeuAlaArGIserAlaIlePheSerValThrTrpProSer 334
   |||
963  CATCTCCACCCCTGGGCCGCTGTCATCTCTCTGTGACATACCCCTCAC 1012
   |||
334  roAspIlePheLeuValIleLysLeuGIuLysValLeuGIuGIuLysP 350
   |||
1013  CTGACATCTTCTCGTCAATGAGTTGGAAGGTGCTTCCACAGAGGGGAC 1062
   |||
351  IleSerGIuCySerGIuProGIyMetValLeuLysGIuValaAspThrAl 367
   |||
1063  ATCAGTGAAGTGTGTGAGCTTACATGGTGTGAAGAAAGTGGACACAGC 1112
   |||
367  aLysAsnLysGIuLysLeuGIuLysLeuArgLeuAlaIaGIuGIuPheC 384
   |||
1113  CAAGAACAAGAGAGAGTAGAAGACTGGCTGGCGCGCGAGCAGTCTCT 1162
   |||
384  yrThrArgLeuGIyArgTrpArgMetProPheAlaTrpThrAlaValHIs 400
   |||
1163  GCACCGCCTGTGGCGCGCTACCCGATGCCCTTCCCTGACAGCGCCCTGC 1212
   |||
401  LeuAlaAsnIleValSerSerAlaGIyGIuLeuAspArgAspSerAsp 417
   |||
1213  TTGGCCAAATCGTAGAGCGCGCTGGGAGCTGGACCGGACTCTGACTC 1262
   |||
417  rGIuGIuLysArgArGIProAlaTrpThrAspArgArGIArgGIyProG 434
   |||
1263  GGAGGGCGAGCGCGCGCGCAGCTGACAGACCGCGCGCTGGGGGCCCC 1312
   |||
434  IAspArgAlaSerSerGIyAspAspAlaCySerPheSerGIyPheArg 450
   |||
1313  AGGACCGGCGAGTAGTGGTGGGAGCAGCGCTGCAGCTTCTGTGGCTTCCGT 1362
   |||
451  ProAlaThrLeuThrValThrAsnPhePheLysGIuGIuAlaGIuArgLe 467
   |||
1363  CCAGCCACCCCTAAGCTGTCAAAACTTCTTTAAGCAGAGGCGTGAAGCACT 1412
   |||
467  uSerAspGIuAspLeuPheLysPheLeuAlaAspMetArgArGIProSers 484
   |||
1413  CAGTGACGAGGACCTTCAAGTCTCGGCTGACATGAGGCCCGCTGTGT 1462
   |||
484  erLeuLeuArGIArgLeuArGIProValThrAlaGIuLeuLysIleAspIle 500
   |||
1463  CCGTGTGGGGGAGCTACGTCTGTGACTGCCAAGCTCAAGATCCACAT 1512
   |||
501  SerProAlaProGIuAsnProHIsPheCyLeuSerProGIuLeuLeuHIs 517
   |||
1513  TCTCGGCGCTCTGAATAATCCCACTTCTGTCTCCCTGAGACTCTTCA 1562
   |||
517  sIleLysProTrpYrProAspProArgGIyArgProThrLysGIuIleLeu 534
   |||
1563  TATCAAGCCCTACCCCGACCCGAGGGCGCGGCCCAAGAGGATTTGG 1612
   |||
534  IuPheProAlaArgGIuValIlyrAlaProHIsThrSerTrpArgAsnLeu 550
   |||
1613  AGTTCCCGCGCGGAGTCTATGCCCCCATACCAAGCTACAGGAACCTG 1662
   |||
551  LeuTrpValTrpProHIsSerLeuAsnPheSerSerArgGIuGIySerVa 567
   |||
1663  CTGTAGCTGAGACCCGACAGCCTCAAACTTCAGCAGCGCGCAGGCTCGCT 1712
   |||
567  IArgAsnLeuAlaValArgValGIuTrpMetThrGIyGIuAspProSerG 584
   |||
1713  GCGCAACCTTGTGCTGCGAGTGCAGTACATGACAGCGGAGGCCACCGCC 1762
   |||
584  IuAlaLeuProValIlePheGIyLysSerSerCySerGIuPheThrArg 600

```

```
|||||
1763 AGGCTTGGCGGTCATCTTTGGCAAGTCCAGCTGCAGTGAATTTACCCG 1812
601 GLUALAphEThrProValValTyrHisAsnLysSerProGluPhETyG1 617
|||||
1813 GAGGCTTACACACCGGTGCTACCATTAACAAGTCCCCGAGTTCTACGA 1862
617 uGluPhelLysLeuHisLeuProAlaCysValThrGluAsnHisLeuL 634
1863 GGAGGTCAAGCTGCATCTTCCAGCTGGCTGCACAGAAACCATCACTGCG 1912
634 euPhEThrPhETyHisValSerCysGlnProArgProGlyThrAlaLeu 650
|||||
1913 TGTTCACCTTCTACCATGTGCAGCTGCAGCCCCGGCGGACCTGCTG 1962
651 GluThrProValGlyPhEThrTrpIleProLeuLeuGlnHisGlyArgLe 667
1963 GAGACACCGGTGGCTTACTTGATCCACAGCTGCAGACAGGCGGCT 2012
667 uArgThrGlyProPhETyCysLeuProValSerValAspGlnProPro 684
2013 GAGGACCGGCGCTTCTGCTCCAGTGTCTGTGACACCGCGCGGCCCA 2062
684 eTyrSerValLeuThrProAspValAlaLeuProGlyMetArgTrpVal 700
2063 GGTATTCGTGCTCACACCCAGTGGCGCTTCCGGGCATGGCGTGGGTG 2112
701 AspGlyHisLysGlyValPheSerValGluLeuThrAlaValSerSerVa 717
2113 GACGGTCAACAAGGCGGTGTCAGTGTGAGCTCACGGCGGTGCTCTGT 2162
717 HisProGlnAspProTyrLeuAspLysPhePhEThrLeuValHisValL 734
2163 GCACCCCAAGACCCCTACCTGAGAAATTTCTACCTGGTGCACGCTCC 2212
734 euGluGluGlyAlaPheProPheArgLeuLysAspThrValLeuSerGlu 750
2213 TGGAGAGAGGAGCTTCCCATTCGGGCTCAAGACACTGTCTGAGCGAG 2262
751 GlyAsnValGluGlnGluLeuArgAlaSerLeuAlaAlaLeuArgLeuAl 767
2263 GGCAACGTGGACAGGAGCTGGGGCCAGTCTTCAGCACACTGGCCCTGGC 2312
767 aSerProGluProLeuValAlaPheSerHisHisValLeuAspLysLeuV 784
2313 CAGCCCCAGAACCCCTGTGGCTTCTCCACACAGCTGCTGCACAACGCTCG 2362
784 AlaArgLeuValIleArgProProIleIleSerGlyGlnIleValAspLeu 800
2363 TGGCTGTGGTCATCAGGCCCGCGATCATCAGTGGCCAGATTGTGAACCTG 2412
801 GlyArgGlyAlaPheGluAlaMetAlaHisValValSerLeuValHisAr 817
2413 GGCGGTGGAGCCTTTGAAGCAATGGCCCATGTAGTCAAGCCTTGTTCAACG 2462
817 gSerLeuGluAlaAlaGlnAspAlaArgGlyHisCysProGlnLeuAlaA 834
2463 GAGCCTGGAGGACCCAGAGATGGCCCGGCTGACATGCCACACAGCTGGCTG 2512
834 IaTyrValHisTyrAlaPheArgLeuProGlyThrGluProSerLeuPro 850
2513 CCTACGTCCATCAGCTTTCGCTTCTGCTGACACTGAGCCAGGCTTCGCG 2562
851 AspGlyAlaProProValThrValGlnAlaAlaThrLeuAlaArgGlySe 867
2563 GATGGGGCCCTCCAGTGCAGTGCAGGCTGGCCACACTGGCGCTGGCTC 2612
867 rGlyArgProAlaSerLeuTyrLeuAlaArgSerLysSerIleSerSers 884
2613 TGTGTGCGCCCAAGCCTTACCTGAGGGCGGTCCACAAGCATCGACGCA 2662
884 eArgAsnProAspLeuAlaValAlaProGlySerValAspAspGluValSer 900
|||||

2663 GCAACCCCTGACCTCGCGGTGGCCCCCTGTGCTGTGATGACGAGGTTTCC 2712
901 ArgIleLeuAlaSerLysLeuLeuHisGluGluLeuAlaLeuGlnTrpVa 917
2713 CGCATCCTGGCGACCAAGCTGTTCACAGAGAGTGTGCTGCAGTGGGT 2762
917 lValSerSerSerAlaValArgGluAlaIleLeuGlnHisAlaTrpPheP 934
2763 GGTAGCAGCAGTCCCTACCGAGGCCATCTCCAGACAGCCCTGGTTCT 2812
934 hePheGlnLeuMetValLysSerMetAlaLeuHisLeuLeuGlyGln 950
2813 TCTTCCAGCTCATGTGTAAGATGGCTGCACCTGCTGCTGGCCAG 2862
951 ArgLeuAspThrProArgLysLeuArgPheProGlyLysPheLeuAspAs 967
2863 CGACTAACACACCCCGCAAGCTGGCTTCCCGGAGCGCTTCTGAGCA 2912
967 PheThrAlaLeuValGlySerValGlyLeuGluValIleThrArgValH 984
2913 CATCAGTGCCTTGGTGGCTCTGTGGGCTGGAGGTCAATCACCCGTGTC 2962
984 lLysAspValGluLeuAlaGluHisLeuAsnAlaSerLeuAlaPhePhe 1000
2963 ACAAGGATGTGGAGCTGGCCGACACCTCAACGCCAGCCGTGGCTTCTTC 3012
1001 LeuSerAspLeuLeuSerLeuValAspArgGlyPheValPheSerLeuVa 1017
3013 CTACGATACCTTCTGTGCTGTGGTGACCGGGCTTGTCTTCAGCTGGGT 3062
1017 lArgAlaHisTyrLysGlnValAlaThrArgLeuGlnSerSerProAspN 1034
3063 CCGGGCCCACTACAGAGAGGTGGCCACGGCGGTCCAGTGTCCCTCAATC 3112
1034 roAlaAlaLeuLeuThrLeuArgMetGluPheThrArgIleLeuCysSer 1050
3113 CAGCAGCCCTGTGACCTTGCATGGAATTCACCGCATCTCTGTGCAGC 3162
1051 HisGluHisTyrValThrLeuAsnLeuProCysCysProLeuSerProP 1067
3163 CACGAGCAGTACGTGACCTCAACCTCCCTGTGCTGCCCTGCACACTGC 3212
1067 oAlaSerProSerProSerValSerSerThrThrSerGlnSerSerThrP 1084
3213 AGCCTGCGCTCCCTCTGTGTCTCCACACACTCCAGAGCTCCACCT 3262
1084 heSerSerGlnAlaProAspProLysValThrSerMetPheGluLeuSer 1100
3263 TCTCCAGCCAGACCCCGGAGCCCAAGAGTGCACGATGTTCCGACTGAGT 3312
1101 GlyProPheArgGlnGlnHisPheLeuAlaGlyLeuLeuLeuThrGluLe 1117
3313 GGACCATTCGCGCAGCAGCACTTCTAGCTGGGCTCTGCTGACGAGCT 3362
1117 uAlaLeuAlaLeuGlnProGluAlaGluGlyAlaPheLeuLeuHisLysL 1134
3363 GGCACGTGGCCCTCGAACCTGAGGCTGAAGGGCATTCCTGTTGCACAA 3412
1134 yAlaIleSerAlaValHisSerLeuLeuCysGlyHisAspThrAspPro 1150
3413 AGGCACTACAGTGTGTGCACAGCTCTATGTGGCCATGACATGACATCCC 3462
1151 ArgTyrAlaGluAlaThrValLysAlaArgValAlaGluLeuTyrLeuPr 1167
3463 CGCTACGCGGAGGCACTGTGAAGCTCTGTGGCGGAGCTGTACTGTGC 3512
1167 oLeuLeuSerIleAlaArgAspThrLeuProArgLeuHisAspPheAlaG 1184
3513 ACTGCTTGTGATTGCAGGGGATACCTTGCCACGGCTGCATGACTTTGCTG 3562
1184 lGluGlyProGlyGlnArgSerArgLeuAlaSerMetLeuAspSerAspThr 1200
3563 AGGGCCCAAGGTCAAGCGGTCAAGACTGGCTCAATGCTTGACTCAACAC 3612
```

1201 GluGlyGluGlyAspIleAlaGlyThrIleAsnProSerValAlaMetAl 1217  
|||||  
3613 GAAGGCGAAGGGGACATGGCGGTACCATCAACCCCTCTGTGGCCATGGC 3662  
1217 aileAlaGlyGlyProLeuAlaProGlySerArgAlaSerIleSerGlnG 1234  
|||||  
3663 CATTGCTGGTGGCCCCCTAGCCCCCTGGCTCCGGGGCCAGCATCTCCCAAG 3712  
1234 1yProProThrAlaSerArgAlaGlyCysAlaLeuSerAlaGlySerSer 1250  
|||||  
3713 GGGCACCACAGGCTCTCGCGCAGGCTGGCCCTCTCTGCTGAGTCAAGC 3762  
1251 ArgThrLeuLeuAlaCysValLeuThrPValLeuLysAsnThrGluProAl 1267  
3763 CGGACCTTGGCTGGCTGTGTGTGGTGTCTGAAAAACACCGAGCGGC 3812  
1267 aLeuLeuGlnArgTrpAlaThrAspLeuThrLeuProGlnLeuGlyArgL 1284  
|||||  
3813 GCTCTCTCAGCGCTGGCGCCACTGACCTGACACCTCCCGACGCTGGAGCTC 3862  
1284 euleuAspLeuLeuTyrLeuCysLeuAlaAlaPheGluTyrTyrGlyLys 1300  
|||||  
3863 TGTGGACTTGTCTTACCTTGGCTGGCTGGCTTGTAGTACAAAGGGGAAA 3912  
1301 LysAlaPheGluArgIleAsnSerLeuThrPheLysLysSerLeuAspPhe 1317  
|||||  
3913 AAGGCTTTTAAACGCATCAACACAGCTCACCATTCAAAAAATCTCTGATAT 3962  
1317 tLysAlaArgLeuGlnGluAlaIleLeuGlyThrIleGlyAlaArgGlnG 1334  
|||||  
3963 GAAGCGCCGCGCTAGAGGAAAGCCATTCTGGGTACCATCGAGCTCGACAAG 4012  
1334 LuMetValArgArgSerArgGluArgSerProPheGlyAsnProGluAsn 1350  
|||||  
4013 AAATGGTTCGCGCAAGTCGTAGAGGAGCCGTTTGGGAATCCAGAGAAC 4062  
1351 ValArgTrpArgLysSerValThrHisTrpLysGlnThrSerAspArgVa 1367  
|||||  
4063 GTGGCGTGGGGGAAGAGCGTCACACACTGGAGAGAAACCTCACAGCCGCT 4112  
1367 LAspLysThrLysAspGluMetGluHisGluAlaLeuValGluGlyAsnL 1384  
|||||  
4113 GGACAAACACCAAGGATGAAATGMAACAGGCGCTTGTGTGAAGGGGAACC 4162  
1384 euAlaThrGlnAlaSerLeuValValLeuAspThrLeuGlnIleIleVal 1400  
|||||  
4163 TGGCAACCGAGGCAAGCTTACTGTTTGGACACACTGGAGATCATCTGTG 4212  
1401 GlnThrValMetLeuSerGluAlaArgGluSerValLeuGlyAlaValLe 1417  
|||||  
4213 CAGACGGTGTGCTTTCAGAAAGCCCGGAGAGGCTTGGGGGACAGTCT 4262  
1417 uLysValValLeuTyrSerLeuGlySerAlaGlnSerAlaLeuPheLeu 1434  
|||||  
4263 GAAGGTTGTGTGTACAGCTCGGGCACTGCCCAAGTGCCTTCTTCTTGC 4312  
1434 LnhISGLYLeuAlaThrGlnArgAlaLeuValSerLysPheProGluLeu 1450  
|||||  
4313 AGCATGGCGCTGGCCACCAAGGGCCCTGTGTGTCAAAGTTCCGGAGCTG 4362  
1451 LeuPheGluGluAspThrGluLeuCysAlaAspLeuCysLeuArgLeuLe 1467  
|||||  
4363 CTGTTCCGAGGAGACAGGAGCTGTGTGCCGACTGTGCTGAGGCTCTCT 4412  
1467 uArgHisCysGlySerArgIleSerThrIleArgThrHisAlaSerLys 1484  
|||||  
4413 ACGACACTGTGGCAGCCGATCAGCACCATCCGACACCGCCAGCCGCTCT 4462  
1484 eLeuTyrLeuLeuMetArgGlnAsnPheGluIleGlyHisAsnPheAla 1500  
|||||  
4463 CGCGTACCTCTCATGTGGACAGAACTTCGAGATCGGCCCAACAACCTTGGC 4512

1501 ArgValLysMetGlnValThrMetSerLeuSerSerLeuValGlyThrTh 1517  
|||||  
4513 CGTGTGAAGATGCAGAGTACCATGTCTCTCTGCTCCCTGGTGGGAGAC 4562  
1517 rGlnAsnPheSerGluGluHisLeuArgArgSerLeuLysThrIleLeuT 1534  
|||||  
4563 GCAGAACTTCAGTGAAGAGACACTGGCAGCTTCACTCAAAACATCTCTCA 4612  
1534 hTyrAlaGluGluAspMetGlyLeuArgAspSerThrPheAlaGluGln 1550  
|||||  
4613 CCTATGCTGAGAGACATGGGGCTGGGACACACACTTTCGACAGACAG 4662  
1551 ValGlnAspLeuMetPheAsnLeuHisMetIleLeuThrAspThrVally 1567  
|||||  
4663 GTCCAGGACCTGATTTCAACCTGACATGATCTCGAGCGACAGGTGAA 4712  
1567 sMetLysGluHisGlnGluAspProGluMetLeuIleAspLeuMetTyr 1584  
|||||  
4713 GATGAAGGAAACACAGAGAGACCTGAGATGCTCATCGACCTCATGTACA 4762  
1584 rGileAlaArgGlyTyrGlnGlySerProAspLeuArgLeuThrTyrPleu 1600  
|||||  
4763 GAATGGCCCGGGCTTACCAAGGCTTACCGGACCTTGGCTGACTGTTG 4812  
1601 GlnAsnMetAlaGlyLysHisAlaGluLeuGluGlyAsnHisAlaGluAla 1617  
|||||  
4813 CAGAACATGGCGCGGAAACACCGGAGCTGGGCAACACCGCGAGCGCGC 4862  
1617 eGlnCysMetValHisAlaAla 1624  
|||||  
4863 CCAGTGCATGTGTGCACGGCGCC 4884  
seq\_name: gb\_pat:AX172874  
seq\_documentation\_block:  
LOCUS AX172874 6828 bp DNA linear PAT 03-JUL-2001  
DEFINITION Sequence 1 from Patent WO0142297.  
ACCESSION AX172874  
VERSION AX172874.1 GI:14597911  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 6828)  
AUTHORS Lu,P., Garman,J.D. and Candia,A.F.  
TITLE Clasp-3 transmembrane protein  
JOURNAL Patent: WO 0142297-A.1 14-JUN-2001;  
Arbor Vita Corporation (US)  
FEATURES  
Source  
1..6828  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="full-length human CLASP-3 cDNA"  
22..6294  
/note="human CLASP-3"  
/codon\_start=1  
/protein\_id="C043726.1"  
/db\_xref="GI:14597912"  
/translation="MAERAFQAKISRTVAAYRQISQYSGSPQLKLNINIVNIS  
HHTTLEAVDVEDLELYLTHPLAVDSGLRDLLEPPDILEVYSPRDCRTLVSA  
VPESEMDPHVRCIRSYPEDMAIVIRKYKILGTGNPMTLDKROKROGLPQVYES  
DEAPDNGSYDDDDDLKRMSISDIDTPRGSMASITPDILNPSDALPLNDRTPREE  
IDKQNDQRSNHRKELFALHPSPDEEPIERLSVDPDIPREHGOGLLYKCSLAKET  
IEBIFASLALYDVKEKKISENFYFDLNSQMKGLLRPHVPAALITTLARSAIFESIT  
YPSQDVLVLKLEKVLQOQDICEAPYMFKEADATKNKLEKLSQADQFCQGLG  
KYRMPFMTAIHLMNIIVSASLSERSTEVEISGRKGSMSERMSIYGRSLERT  
TSGDDACNLTSFSPATLTVTNPFKQSGDRSDLDLKKPLADMRSPSYLRRLPITAO  
LKIDISAPNPNPHCLTPELLQVLYKLPDSDVRRTRRELEPRADVYVPMPTVKNLLYI  
YPSIDIANRQGSARNITVAVQFMYGDPDSNAPVIFGSSCESFEKATVAVYHNR  
SPDFHEIKVLPATLTDHNLHLEFYHNSQDKQKPTLETGVYTIMPLQNGRLKT  
GQFLPVSLEKPRDAYSVLSEVPLPGMKVNDHKGVFNVENVAVSSITQDPYLDKF  
FALYNALDEHLFPRIQDMRIMENINLENELKSSISALNSQLEPVPVYRFLHLDLKL







```

|||||.....
5001 GATTGATCTAAATGTACAGAAATGGCCAGGGTTACACAGACTCTCCAGAA. 5049
1595 eaurgleuthrtrpleuglnasnmetalaaglylshislaagluengly 1611
5050 .CGATTGACCTGGTTCGAGAACATGGCAGCAGCAGCCTCAGACAGAAAC 5097
1612 AsnHISlaaGluAlaaglnCysmetValHISlaaAlaAlaLeuValAl 1628
5098 AATCAATGCTGAACCTGCACAGTGTCTAGTCCACTCAGCAGCAGCTGTTGC 5147
1628 agluTyrlaAlaLeuengluaspGlnarGHisleuProValIGlyCysV 1645
5148 TGAATATTAGCATCTGTGAGAGACCGGAAATATCTTCTGTGGAGATGG 5197
1645 aIserepHeGlnaenIIeserSerAsnValleuGluGluSerAlaIleser 1661
5198 TAACATTTTCAGAAATATTCTAATGTTTATGAAAGAAATCTGGGCTCTCA 5247
1662 AspaSrpIleuSerProaspGluGluGlyPheCysSerGlyLysHisph 1678
5248 GATGATGTGGTATCTCCAGATGAGAGAGATATCTGCTCTGGAATAACTT 5297
1678 eThnGluLeuGlyLeuValGlyLeuengluGlnAlaAlaaglyTyrrPheT 1695
5298 TACTGAGTCCAGACTGTGGGATTTACTGGAAACAGCAGCTTCTCTCT 5347
1695 hrnetGlyLeuTyrgluAlaValasnguValTyrlYrsasnleuIle 1711
5348 CTATGGCTGCATGTATGAAAGCAGTTAATGAAAGTTATACAAAGACTTATT 5397
1712 ProIleuengluAlaHisArgaspTyrlYsLysLeuAlaAlaValHisgl 1728
5398 CCTATTATAGACTTAATCGGATGCAAGAAACTAATCCACATTCATCG 5447
1728 yLysLeuGlnGluAlaPheThrLysIleMetHisGlnSerSerGlyTrrpG 1745
5448 TAAACTTCAGAAAGCATTTCAGCAAAATTTGTTCAATCAGACTGCTGGG 5497
1745 luarGValIlePheGlyThrTyrrPheArgValGlyPheTyrgluAlaHisph 1761
5498 AGCGGATTTGGCACCATTATTCGTGTGTTTATGGAACCAAGCTTC 5547
1762 GlYaspLeuaspGluGlnGluPheValTyrlYsGluProSerIleThrIly 1778
5548 GGGGATTTGGATGTAACAAGAAATTTGTTTACAGAGCCCTCAATTAACCA 5597
1778 sLeuAlaGluIleSerHisArgleuGluGluPheTyrrGluArgPheG 1795
5598 ACTTGACAGATATCTCAGAGATGGAGGATTTTACGGAGAAAGATTTG 5647
1795 lyaspPvalValGluIleIleLysaspSerTyrrProValaspLysSer 1811
5648 GAGAGGATGTGTGAGAGTATCAAAACTTAATCTCTGAGACAAAGTGT 5697
1812 LysLeuaspSerGlnLysAlaTyrlIleGlnIleThrTyrrValGluProTy 1828
5698 AAATTAAGTCTTAACAAGCATATATTCAGATTACTATGGGAGCCATA 5747
1828 rPheaspThrTyrgluLeuLysaspArgValIleThrTyrrPheaspArgsnT 1845
5748 CTTTGACACATATGATGAGAGACAGATACCTATTTGCGACAAAAATT 5797
1845 yrglyLeuArgThrPheleuPheCysThrProPheThrProaspGlyArg 1861
5798 ACAATCTTCGTCGATCTGACTGATACCCCTTACTTTAGATGGCCGT 5847
1862 AlaHisGlyGluLeuProGluGlnHisLysArgLysThrleuLeuSerTh 1878
5848 GCCCATGGGGAACCTTCAACAATTCAAAAGAAAGCAATTTGCACTAC 5897
1878 rAspHisAlaPheProTyrlIleLysThrArgIleArgValCysHisArgG 1895
|||||.....

```

```

5898 GTCATGACCTTCCCTTATATAAACAAGGTCATGCTCATTAAG 5947
1895 luGluThrValleuThrProValGluValAlaIleGluaspmetGlnLys 1911
5948 AAGAGATCATCTTAAACCAATTAAGTGTCTATTGAGCAACAGCAAAA 5997
1912 LysThrArgGluLeuAlaPheAlaThrGluGlnaspProProaspAlaTy 1928
5998 AAGACACAGAGATTGGCATTTGCAACACATCAGATCCCGAGACCCCAA 6047
1928 smetleuGlnMetValleuenglySerValGlyProThrValasnGln 1945
6048 AATGCTTCAGATGTGACTCCAGGATCTGTAGGCACCAAGTGAATAGG 6097
1945 lYProleuGluValAlaGlnValPheleuAlaGluIleProGluaspPro 1961
6098 GCCCTTTGGAAGTTGCCAGCTTTTCTGTGAATAATCAGTAGACCCA 6147
1962 LysleuPheArgHisHisAsnLysleuArgleuCysPheLysaspPheCy 1978
6148 AAGCTCTTCAGACATCATTAATAACTGCGACTGCTTTTAAAGATTTTAC 6197
1978 slYsLysCysGluaspAlaLeuArgLysAsnLysAlaLeuIleGlyProA 1995
6198 TAAAGGtGTGAAGATGCTTAAGAAATAATAGACTTAATTGGCCCG 6247
1995 spGlnLysGluTyrrHisArgGluLeuGluArgYrAsnTyrrCysArgLeuAr 2011
6248 TTCAAAAGAGATATCAAAAGGAATGG...GGGAACACTTCTTCCCTTAA 6294
2011 gGluAlaLeuGln.ProleuLeuThrGlnArgleuProGluLeuMetAla 2027
6295 AGAGGCCCTTACAGCCCT...AGATCACAGAAAGTCCCTCAGTTATCCAG 6341
2028 ProThr.....ProProGlyLeuArgAsnSerleuAsnArgAlaSerPh 2042
6342 CCAGTATTGCTTGTCCCTGCCACAGACATTCCTTCAATCGAATGAGCTT 6391
2042 eArgLysAlaaspLeu 2047
6392 TCGCAAAATGATGATCTC 6407

seq_name: gb_pat:AX174569
seq_documentation_block:
LOCUS AX174569 7215 bp DNA linear PAT 04-JUL-2001
DEFINITION Sequence 1 from Patent WO0142296.
ACCESSION AX174569
VERSION AX174569.1 GI:14598200
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 7215)
AUTHORS Lu,P., Garman,J.D. and Candia,A.F.
TITLE Clasp-5 transmembrane protein
Patent: WO 0142296-A 1 14-JUN-2001;
JOURNAL Arbor Vita Corporation (US)
FEATURES
Location/Qualifiers
1..7215
1..7215
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="full length human Clasp-5 cDNA"
112..6159
/note="human Clasp-5"
/codon_start=1
/protein_id="CAC43733.1"
/db_xref="GI:14598201"
/translation="MTLNSLDVQLAQLGELDFDDDDLDVFTFKKERTLOPSLPEGY
ELDPHVDVCVQTYIREWLIVNRKNQGSPEICGRKKGSRKDFHKLTKPOTFESYLEC
SEPAAGPRLNVLNLCVSGKGPVTACDFRLSDLPDRLENDLQOVSADEFKQNEE
ARKNRQAELEFALYPSVDEEDAVIEIRVPCREHIGNRILVLTILKFEIETEPLEFA

```





```
111 ..... |||:||||: ::|||:
1257 CTCCTGAGGAAATGGGTTGGATCCAACTCAAAACCTCCACTGTA 1306
455 hrValThrPhePheLeuGlnGluAlaGluSerAspGluAsp 471
   ::|||:|||||:|||||:|||||:|||||:|||||:|||||:
1307 GCGTTAGAGCTTTTCAAGCAGGAGAGATGCCCTTAGCGAGGAGAC 1356
472 LeuPheLeuPheLeuAlaAspMetArgProSerSerLeuLeuArg 488
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1357 TTATTCAAGTTTATAGCTAGACTACAAAAGATCATCTTACGAGACG 1406
488 GLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAlaPro 505
   ::|||:||||:||||:||||:||||:||||:||||:
1407 AGTCAGATCATTCAGGCTTGCTAGACGAGATTTCTACAGCTCCAG 1456
505 LuAsnProHisPheCysLeuSerProGluLeuLeuHisIleLysProTyr 521
   |||:||||:||||:||||:||||:||||:||||:
1457 AGATCATCAATTCGTGCTGACTCTGTAAACATGCTGCCCGTGAACCTTT 1506
522 ProAspProArgGlyArgProThrLysGluIleLeuGluPheProAlaAr 538
   |||:||||:||||:||||:||||:||||:||||:
1507 CCTGAAACCGGACACGCCGCCACAAAGAGATTGTGGAATTTCCACACG 1556
538 GGlValThrAlaProHisThrSerTyrArgAsnLeuLeuTyrValTyrP 555
   |||:||||:||||:||||:||||:||||:||||:
1557 AGAAGTATATGTCCTCACACTGTGTACAGAAACCTTCTATGTCTACC 1606
555 roHisSerLeuAsnPheSerSerArgGlnGlySerValArgAsnLeuAla 571
   |||:||||:||||:||||:||||:||||:||||:
1607 CACGAGGCTCACTTGTGTAACAACTAGCATACGCCGGAACATTACA 1656
572 ValArgValGlnTyrMetThrGlyLysAspProSerGlnAlaLeuProVa 588
   |||:||||:||||:||||:||||:||||:||||:
1657 ATAAAGATCCAGTTTATGTGTGAGAGAGATGCTAGCAATGGAGTCCGGT 1706
588 ILePheGlyLysSerSerCysSerGluPheThrArgGluAlaPheThrP 605
   |||:||||:||||:||||:||||:||||:||||:
1707 CATCTTTGGAAATCCACGCGGCTGAAATTTCTGCAGGAAGTGTACACAG 1756
605 roValValTyrHisAsnLysSerProGluPheTyrGlnGluPheLeu 621
   |||:||||:||||:||||:||||:||||:||||:
1757 CTGTTACTACCATTAATAGCTCTCTGACTTTTATGAAGAAGTAAATTT 1806
622 HisLeuProAlaCysValThrGluAsnHisLeuLeuPheThrPheTy 638
   |||:||||:||||:||||:||||:||||:||||:
1807 AAGCTCCCGGCTAGCTACAGTAATCACACCTCTGTTCACTTCTTA 1856
638 rHisValSerCysGlnProArgProGlyThrAlaLeuGluThrProValG 655
   |||:||||:||||:||||:||||:||||:||||:
1857 CCATATCAGCTGTACGAGAGAGAGGCTCCGTGGAACCTCTCTGTGG 1906
655 LysPheThrTrpIleProLeuLeuGlnIleGlyArgLeuArgThrGlyPro 671
   |||:||||:||||:||||:||||:||||:||||:
1907 GATATCTAGGCTGCCAATTTCTTTAAATGACGTCTTCAACTGAGTCC 1956
672 PheCysLeuProValSerValAspGlnProProProSerTyrSerValLe 688
   |||:||||:||||:||||:||||:||||:||||:
1957 TACGTGTCCCAAGTTCGCTGGAATAATGGCCACCAACTACATCCATGCA 2006
688 uThrProAsp.....ValAlaLeuProGlyMetArgTrpValAspG 702
   ::|||:||||:||||:||||:||||:||||:
2007 TTCGTGAGAGAAAGTCCCATTAACAGAAATCCCTCCATTAAAGTGGGCTGAG 2056
702 LysHisLysGlyValPheSerValGluLeuThrAlaValSerSerValHis 718
   |||:||||:||||:||||:||||:||||:||||:
2057 GACATTAAGGAGTATTTAATATTAAGTGCACAGCTGTTTCTTGTAACAC 2106
719 ProGlnAspProTyrLeuAspLysPhePheThrLeuValHisValLeuG 735
   |||:||||:||||:||||:||||:||||:||||:
2107 ACCCAGGACAAACACCTGAGAAAGTTCTTCACTTGCACATCCCTCGGA 2156
735 u...GluGlyAlaPheProPheArgLeuLysAspThrValLeuSerGluG 751
   |||:||||:||||:||||:||||:||||:||||:
2157 GAGCCAGGTGACTTCCCATCCGCGTGTGATCAGAAATCAGCAGGA 2206
751 LysAsnValGlnGlnGluLeuArgAlaSerLeuAlaLeuLeuArgLeuAla 767
   ::|||:||||:||||:||||:||||:||||:||||:
2207 TGGCGCTGAGCATAGACTGAAAGCTCAGCATCATCTGCTGAATCTCTCC 2256
768 SerProGluProLeuValAlaPheSerHisValLeuAspLysLeuVal 784
   |||:||||:||||:||||:||||:||||:||||:
2257 CGCGCTGAGCCGCGTGTCTCTGTCGACCTGGTGGTGGACAACTCTT 2306
784 LArgLeuValIleArgProProIleIleSerGlyGlnIleValAsnLeuG 801
   ::|||:||||:||||:||||:||||:||||:||||:
2307 CCAGCTCCGCGCACCCCATGTCATCGTGGCAGACAGCAACCTTCT 2356
801 LArgGlyValAlaPheGluAlaMetAlaHisValIleValSerLeuValHisArg 817
   |||:||||:||||:||||:||||:||||:||||:
2357 CCAGTTCGCTTCCGAGTCCGTGTGGTGGCCATGCCAACAGTGTGCACAC 2406
818 SerLeuGluAlaAlaGlnAspAlaArgLysHisCysProGlnLeuAlaAl 834
   |||:||||:||||:||||:||||:||||:||||:
2407 AGCAAGAGACCTGAGCAGAGCAGCATGGAGAGAACTGCTGCTGCTTC 2456
834 aTyrValHisTyrAlaPheArgLeuProGlyThrGluProSerLeuPro. 850
   ::|||:||||:||||:||||:||||:||||:||||:
2457 CTACGTCACCTACGTCTTCCGCTGCGCAGAGGTGCAGAAAGGATGTGCCCA 2506
851 ..AspGlyAlaProProValThrValGlnAlaAla.....ThrLeu 863
   ::|||:||||:||||:||||:||||:||||:||||:
2507 AGTCAGGCGCTCCACCTGCCCTTAGACCTTCGGAGCTACACACAGTAT 2556
864 AlaArgGlySerGlyArgProAlaSerLeuTyrLeuAlaArgSerLysSe 880
   ::|||:||||:||||:||||:||||:||||:||||:
2557 GGGCGCATCATCAGCTGCTGTGAGTTCAAGCTGTGTCAGCGCCGGGT 2606
880 rIleSerSerSerAsnProAspLeuAlaValAlaProGlySerValAspA 897
2607 GATGACACACAGTAACCCAGACCTCGCGGGGACACACTCCGACGACAGG 2656
897 spGluValSerArgIleLeuAlaSerLysLeu..... 907
2657 AGSAAATGAAGAAACATCATGCTTCAAAAGATCGCGGATCGCAATGCAAGC 2706
907 ..... 907
2707 CGAATGCTTACTATGCTCTGTGACAGTAGTAGTCCAAAGTTCACTGC 2756
908 ..... LeuHisGlnGluLeuAlaLeuG 915
2757 AGCCCCAAGGCCAGCCGCAAAAAGCATTTCCATGAGGAGCTTGCCCTTC 2806
915 InThrValIleSerSerSerAlaValArgGluAlaIleLeuGlnHisAla 931
   |||:||||:||||:||||:||||:||||:||||:
2807 ACATGCTGTGTCAGCACCGGA..... 2826
932 TrpPhePhePheGlnLeuMetValLysSerMetAlaLeuHisLeuLeu 948
2827 ..... ATGCTGAAGAGCATGCGCCACAGCATGACATTA 2858
948 uGlyGlnArgLeuAspThrProArgLysLeuArgPheProGlyArgPheL 965
2859 CATGACAAACGCGACAGTTTTCGGAGACGCTGTTTTTCTGACCTTTTCA 2908
965 euAspAspIleThrAlaLeuValGlySerValGlyLeuGluVal..... 979
2909 TGGATACACTAATCTACTATTGTTAATGTGTGCTACCTCGGAATTTGCACCC 2958
980 ..IleThrArgValHisLysAspValGluLeuAlaGlnHisLeuAsnAl 995
2959 CTTTATGTAACACACAGAAAGAAATGAAACAGCGGAAAGATGAAACAT 3008
995 aSerLeuAlaPhePheLeuSerAspLeuLeuSerLeuValAspArgGlyP 1012
3009 CAGCGTGGCTTCTCTGTATGACCTTCTCTCCCTCATCGATCGGGCT 3058
```

1012 heValPheSerLeuValArgAlaHisTyrLysGlnValAlaIleThrArgLeu 1028  
|||||.....| | | | | : : : : : | | | | | : : : : : | | | | |  
3059 TTGGTTTAACCTCATGACATATTATTCAGCCAGCTGTCAGCCAAGCTC 3108  
1029 GlnSerProAsnProAlaAlaLeuThrLeuArgMetGluPheTh 1045  
: : : : : | | | | | : : : : : | | | | | : : : : : | | | | |  
3109 AGTAACCTTCCA.....ACGCTCATTTCCATGAGGCTTAAGTTCCCT 3149  
1045 rArgLLeuCySerHisGlnHisTyrValThrLeuAsnLeuProCysC 1062  
|||||.....| | | | | : : : : : | | | | | : : : : : | | | | |  
3150 GAGAAATCCTGTGATGCCATGACATTAACCTCAATCGAACTTTTCTTA 3199  
1062 ySProLeu...SerProAlaSerProSerProSerValSerSerThr 1077  
3200 TGAATGCTGATACCTGCTCAACATCTCTCTGCTTCATATCT..... 3243  
1078 ThrSerGlnSerSerThrPheSerSerGlnAlaProAspProLysValTh 1094  
|||||.....| | | | | : : : : : | | | | | : : : : : | | | | |  
3244 ...TCCCAAGAACTCAAGCTCCTGCTCCAGCTCCAGGACCAAGAAATCGC 3290  
1094 rSerMetPheGluLeuSerGlyProPheArgGlnGlnHisPheLeuAlaG 1111  
|||||.....| | | | | : : : : : | | | | | : : : : : | | | | |  
3291 CAGCATGTTGATCTGACTTCGAGTAACCCGACAGCACTTCTCACC 3340  
1111 LLeuLeuLeuThrGluLeuAlaLeuAlaLeuGluProGluAlaGluGly 1127  
|||||.....| | | | | : : : : : | | | | | : : : : : | | | | |  
3341 GGCTCCCTTTCACAGAACTGGCTGCTGCTGCTGATGCCGAGGGAAGGA 3390  
1128 AlaPheLeuHisLysLysAlaIleSerAlaValHisSerLeuLeu 1144  
: : : : : | | | | | : : : : : | | | | | : : : : : | | | | |  
3391 ATCAGCAAACTACAAGGAAGCTGTCAGTCAATTCACAGCTGCTAAG 3440  
1144 sGlyHisAspThrAspProArgTyrAlaGluAlaThrValLysAlaArgY 1161  
|||||.....| | | | | : : : : : | | | | | : : : : : | | | | |  
3441 TTCTCAGCACTGAGCCACGCTGCTGCAAAACCAAGAGTGAAAGTCAAAA 3490  
1161 alaAlaGluLeuTyrLeuProLeuLeuSerIleAlaIleArgAspThrLeuPro 1177  
: : : : : | | | | | : : : : : | | | | | : : : : : | | | | |  
3491 TCGGCGCCCTTTACATACCTTTAGTTGGATCATTTTGGATGCTTTGCCA 3540  
1178 ArgLeuHisAspPheAlaGluGlyProGlyGlnArgSerArgLeuAlaG 1194  
: : : : : | | | | | : : : : : | | | | | : : : : : | | | | |  
3541 CAGCTCTGTACCTTACAGTTGCAGATACCTGCAGATACCGCACAGTGG 3590  
1194 rMetLeuAspSerAspThrGluGlyGluGlyAspIleAlaGlyThrIleA 1211  
|||||.....| | | | | : : : : : | | | | | : : : : : | | | | |  
3591 C.....TCGGATGAAGAACAAAGAGA.....GCCGATGCCATTA 3625  
1211 snProSerValAlaMetAlaIleAlaGlyLProLeuAlaProGlySer 1227  
| | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |  
3626 ACCAGAAATGGCTGTGGCCATAGCAGGAAATATTCAATTGAAAACA 3675  
1228 ArgAlaSerIleSerGlnGlyProProThrAlaSerArgAlaGlyCysAl 1244  
: : : : : | | | | | : : : : : | | | | | : : : : : | | | | |  
3676 AGTGAATAGTCTGTCTTCCTGCC.....TATAGACAGTAACAAT 3719  
1244 alaUserAlaGluSerSerArgThrLeuLeuAlaCysValLeuThrVal 1261  
|||||.....| | | | | : : : : : | | | | | : : : : : | | | | |  
3720 GTGAACGGCGGACATCTGCAACCTGATGATCTCTTCCTGGATCA 3769  
1261 euLysAsnThrGluProAlaLeuLeuGlnArgThrAlaThrAspLeuThr 1277  
: : : : : | | | | | : : : : : | | | | | : : : : : | | | | |  
3770 TGAATAATGCTGATTCAGAGCTCATTAAGAGGATTCGTGCTGCCCA 3819  
1278 LeuProGlnLeuGlyArgLeuLeuAspLeuLeuTyrLeuCysLeuAla 1294  
|||||.....| | | | | : : : : : | | | | | : : : : : | | | | |  
3820 TCAACGACAGCTCAACAGATTTTAAATCTACTTTTCATCTGTGTTATG 3869  
1294 aRheGluTyrLysGlyLysLysAlaPheGluArgIleAsnSerLeuThrP 1311  
|||||.....| | | | | : : : : : | | | | | : : : : : | | | | |  
3870 TTTTGAATTAAGGTAACAGAGTTCTGACAAAGTCACTACCCCAAGTCC 3919

1311 heLysLysSerLeuAspMetLysAlaArgLeuGluGlnAlaIleLeuGly 1327  
: : : : : | | | | | : : : : : | | | | | : : : : : | | | | |  
3920 TGCAGAAATCAAGGATGTCAAGGCCGCTGAGAGAGCTTGTGCTGGT 3969  
1328 ThrIleGlyAlaArgGlnGluMetValArgArg.....SerArgG 1341  
|||||.....| | | | | : : : : : | | | | | : : : : : | | | | |  
3970 GGGGAAGGGGGCCAGAGGAGATGATGGCCGCCGCGCTCCAGGGAACA 4019  
1341 uArgSerProPheGlyAsnProGluAsnValArgTrpArgLysSerValT 1358  
| | | | | | | | | | | | | | | : : : : : | | | | | : : : : : | | | | |  
4020 CCGATTTCGA...GGCCTAAATGAATAATTGAGATGAGAAGAAAGCAGA 4066  
1358 hrHisTyrLysGlnThrSerAspArgValAspLysThrLysAspGluMet 1374  
|||||.....| | | | | : : : : : | | | | | : : : : : | | | | |  
4067 CACATTGGCGGCAAGCTAATGAGAAGCTAATGAATAAACAAGCCAGTTA 4116  
1375 GlnHisGluAlaLeuValGluLysAsnLeuAlaThrGluAlaSerLeuVal 1391  
: : : : : | | | | | : : : : : | | | | | : : : : : | | | | |  
4117 GATCAAGAAAGCTTGATCAGTGGCAATCTGGCTACAGAAACACATTTAAT 4166  
1391 lValLeuAspThrLeuGluIleIleValGlnThrValMetLeuSerGluA 1408  
: : : : : | | | | | : : : : : | | | | | : : : : : | | | | |  
4167 CATCCGTGATATGACAGAAACATATATCCAGGCGAGCTCGGCTGAGACT 4216  
1408 lArgGluSerValLeuGlyAlaValLeuLysValValLeuTyrSerLeu 1424  
: : : : : | | | | | : : : : : | | | | | : : : : : | | | | |  
4217 GTAAAGACAGCTGCTGGAGGCTGTGAGAGGTCTGTGAATATCTGTG 4266  
1425 GlySerAlaGlnSerAlaLeuPheLeuGlnHisGlyLeuAlaThrGlnAr 1441  
: : : : : | | | | | : : : : : | | | | | : : : : : | | | | |  
4267 AACGTGATACAGATGATACCAACCTGACTGACTGCTTGGCAACCTCGG 4316  
1441 gAlaLeuValSerLysPheProGlnLeuLeuPheGluGluAspThrGluL 1458  
|||||.....| | | | | : : : : : | | | | | : : : : : | | | | |  
4317 TCCTCTCAATCCGCAAGTTTGGAGACTTACTCTTCGAAAGAGGGGGAAC 4366  
1458 euCysAlaAspLeuCysLeuArgLeuArgHisCysGlySerArgIle 1474  
| | | | | | | | | | | | | | | : : : : : | | | | | : : : : : | | | | |  
4367 AGTGTTCGACCTATGTACCAACAGTCTGCACACTGACGACGACATG 4416  
1475 SerThrIleArgThrHisAlaSerAlaSerLeuTyrLeuMetArgG 1491  
: : : : : | | | | | : : : : : | | | | | : : : : : | | | | |  
4417 GATGTACCCGGAGCCAGACCTGTGCACCTTATACCTCTCATAGAGTT 4466  
1491 nasPheGluIleGlyHisAsnPheAlaArgValLysMetGlnValThr 1508  
: : : : : | | | | | : : : : : | | | | | : : : : : | | | | |  
4467 CAGTTTGGAGCCACCACTAATTTTGCAGAGTAAGATGCAAGTAACA 4516  
1508 eTSerLeuSerSerLeuValGlyThrThrGlnAsnPheSerGluGlnHis 1524  
| | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |  
4517 TGTCCCTGGCATCTTTGGTGGAGAGACACCAAGCTTTAATGAAGAAGC 4566  
1525 LeuArgArgSerLeuLysThrIleLeuThrTyrAlaGluGlnAspMetG 1541  
|||||.....| | | | | : : : : : | | | | | : : : : : | | | | |  
4567 CTGAAGAAGTCTTGAGACAATTTTGGCTATTCAGAAAGACACAGC 4616  
1541 yLeuArgAspSerThrPheAlaGluGlnValGlnAspLeuMetPheAsn 1558  
: : : : : | | | | | : : : : : | | | | | : : : : : | | | | |  
4617 CATGCAGATGACTCTCTTCCACCCAGGTGAGAGAACTCTGTATATC 4666  
1558 euHisMetIleLeuThrAspThrValLysMetLysGlnHisGlnGluAsp 1574  
| | | | | | | | | | | | | | | : : : : : | | | | | : : : : : | | | | |  
4667 TGAATAGCATCTTATATGACAGAGTGAAGAAATGAGGAATTTTCAGAGAT 4716  
1575 ProGluMetLeuLeuAspLeuMetLysArgIleAlaArgGlyTyrGlnG 1591  
|||||.....| | | | | : : : : : | | | | | : : : : : | | | | |  
4717 CCGTAGATGCTTATGAGATCATGTACAGATTTGCCAAGAAATTCACAGCC 4766  
1591 ySerProAspLeuArgLeuThrTrpLeuGlnAsnMetAlaGlyLysHisA 1608  
|||||.....| | | | | : : : : : | | | | | : : : : : | | | | |  
4767 ATCTCTGATCTGGCGGTGACCTGCTCCAGAAACATGAGCAAGAAACA 4816  
1608 lArgLeuGlyAsnHisAlaGluAlaIleGlnCysMetValHisAlaAla 1624

```

4817 CCAAGAAGAGTGTCTACACGAGGCTGCATGTGCTGGTGACGCCGCT 4866
1625 AAlaValAlaGluTyrLeuAlaLeuLeuGluAspGlnArgHisLeuP 1641
4867 GCGTTAGTGGCTGATATCTGACATGCTGGAGACCACACGCTACTGCC 4916
1641 oValGlyCysValSerPheGlnAsnIleSerSerAsnValLeuGluGln 1658
4917 CGTGGGAGGTGACGCTCCAGAAATTTCTTCATGTGCTGGAGAGT 4966
1658 eAlaIleSerAspAspIleLeuSerProAspGluGluGlyPheCysSer 1674
4967 CTGTGGCTCTGAGAGACACCTGTCACTGACGAGGATGGGGTGTCCGA 5016
1675 GlyHisHisPheThrGluLeuGlyLeuValGlyLeuGluGlnAla 1691
5017 GGGCAGTACTTCACCGAGAGTGGCTGTGGCTCTCTGGAGCAGGCCG 5066
1691 aGlyTyrPheThrMetGlyGlyLeuTyrGluAlaValAsnGluValTyr 1708
5067 GGAGCTCTTCAGCAGCGAGGCTTATATGACAGCTTAATGAGGTCTACA 5116
1708 yAsnLeuIleProIleLeuGluAlaHisArgAspTyrIleCysLeuAla 1724
5117 AGCTGGTCATCCCATCTTAGAAGCGCATGAGAAATTCGGAGAGCTGACA 5166
1725 AAlaValIleGlyLeuLeuGlnGluAlaPheThrLysIleMetHisGln 1741
5167 CTCACATCACAGCAAGCTGCAGAGAGCTTCGACAGCATGCTTACACAGA 5216
1741 rSerGlyTyrGluArgValPheGlyThrTyrPheArgValGlyPheTyrG 1758
5217 TCAT.....AAGAATGTTTGAACCTACTTCGAGTTGGTTCTTGG 5260
1758 yAlaHisPheGlyAspLeuAspGluGlnGluPheValTyrIleGluPro 1774
5261 GATTCCAAATTTGGGATTTGGATGAACAGAGTTGTCTACAAAGAGCCT 5310
1775 SerIleThrLysLeuAlaGluIleSerHisArgLeuGluGluPheTyr 1791
5311 GCATTTACCAAGCTCTCTGATCTCACATAGCTAGAGCATTTTATG 5360
1791 rGluArgPheGlyAspAspValValGluIleIleLysAspSerTyrPro 1808
5361 TCATGTTTGGTGCAGAAATTTGTGAGAGATTAAGATTCACACTCCG 5410
1808 aAlaPylSerLysLeuAspSerGlnLysAlaTyrIleGlnIleThrTyr 1824
5411 TGGCAAAACCAAGTTGGATCTTACACAGGCCCTTACATACGATCCTTT 5460
1825 ValGluProTyrPheAspThrTyrGluLeuLysAspArgValThrTyrP 1841
5461 GTGGAGCCCTACTTGTATGATGATGAGATGAAGACAGGCTCATACTT 5510
1841 eAspArgAsnTyrGlyLeuArgThrPheLeuPheCysThrProPheThr 1858
5511 TGGAAATAATTTCAACCTCCGAGGTTCATGATCAACACCCCTTCACCC 5560
1858 roAspGlyArgAlaHisGlyLeuLeuProGluGlnHisLysArgLysThr 1874
5561 TGGAGGGGCGGCTCGGGAGAGCTGCATGAGCAGTACAGACAGATCAGCG 5610
1875 LeuLeuSerThrAspHisAlaPheProTyrIleLysThrArgIleArgVa 1891
5611 GTCTGACACCTATGACGCGCTTCCCTACATCAGACAGCAGATCAGCG 5660
1891 LcysHisArgGluGluThrValLeuThrProValGluValAlaIleGluA 1908
5661 CATCCAAAGAGAGAGTTGTGTTTGAACCGATTAAGATTGCCATTGAAG 5710
1908 spMetGlnLysLysThrArgGluLeuAlaPheAlaThrGluGlnAspPro 1924
|||||..... ||| .....|||

```

```

5711 ACATGAGAGAGAGACCTCGAGTTAGCATTTGCCATTTACAGGAGCCG 5760
1925 ProAspAlaLysMetLeuGlnMetValLeuGlnGlySerValGlyProTh 1941
5761 CCTGATGCAAGATGCTTACATGTGTGCTGCAAGGCTTGTGGAGCTAC 5810
1941 rValAsnGlnGlyProLeuGluValAlaGlnValPheLeuAlaGluIleP 1958
5811 TGTAAATCAGGAGCACCTGGAAGTAGCCCAAGTGTGTTTGGCTGAATTC 5860
1958 roGluAspProLysLeuPheArgHisHisAsnLysLeuArgLeuCysPhe 1974
5861 CTGTGATTCCAAACCTCTATGACATCACAAACAAGTTGAGGTATGCTTT 5910
1975 LysAspPheCysLysLysCysGluAspAlaLeuArgLysAsnLysAlaLe 1991
5911 AAGGAATTCATCATGAGATGTGTGTCGAACCTTACAGAAAAACAAGCCT 5960
1991 uIleGlyProAspGlnLysGluTyrHisArgGluLeuGluArgAsnTyrC 2008
5961 CATCAGCGCAGACAGCAGCAATATACAGCAAGACTCAAAAGAATATA 6010
2008 yArgLeuArgGluAlaLeuGlnProLeuLeuThrGlnArgLeuProGln 2024
6011 ACAACCTAAAGAGAACTCAGGCCAATGATCGAGCGGAAATTCACAGA 6060
2025 LeuMetAlaPro.....ThrProGlyLeuArgAsnSerLys 2038
6061 CTGTCAAGCCCAATATTCAGAGTTGAGAGTCAAAAGAGGAGCTCTTCCA 6110
2038 nArgAlaSerPheArgLysAlaAsp 2046
6111 CAGATCTAGTTTCAGGAATGTGAA 6135

seq_name: gb_ov:AY016023

seq_documentation block:
LOCUS AY016023 167215 bp DNA linear VRT 01-FEB-2002
DEFINITION Spherooides nephelus alpha globin gene cluster, complete sequence.
ACCESSION AY016023
VERSION AY016023.1 GI:18463963
KEYWORDS
SOURCE
ORGANISM
Spherooides nephelus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Spherooides.
1 (bases 1 to 167215)
REFERENCE
1 Flint,J., Tufarelli,C., Peden,J., Clark,K., Daniels,R.J.,
Hardison,R., Miller,W., Phillipsen,S., Tan-Un,K.C., McMorrow,T.,
Frampton,J., Alter,B.P., Frischauf,A.M. and Higgs,D.R.
Comparative genome analysis delimits a chromosomal domain and
identifies key regulatory elements in the alpha globin cluster
Hum. Mol. Genet. 10 (4), 371-382 (2001)
JOURNAL
MEDLINE 21096913
PUBMED 11157800
REFERENCE
1 (bases 1 to 167215)
Flint,J., Tufarelli,C., Peden,J., Clark,K., Daniels,R.J.,
Hardison,R., Miller,W., Phillipsen,S., Tan-Un,K.C., McMorrow,T.,
Frampton,J., Alter,B.P., Frischauf,A.M. and Higgs,D.R.
Direct Submission
Submitted (06-DEC-2000) MRC Molecular Haematology Unit, Institute
of Molecular Medicine, John Radcliffe Hospital, Oxford, Oxon OX3
9DS, UK
FEATURES
source
1..167215
/organism="Spherooides nephelus"
/db_xref="taxon:39110"
1510..1530
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=c

```

```
repeat_region 2315..2350
/note="C-rich; low complexity"
repeat_region 2643..2663
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=ca
repeat_region 3620..3676
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=tcg
repeat_region 4231..4284
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=ca
repeat_region 4638..4678
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=tcg
repeat_region 5566..5600
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=ca
repeat_region 5840..5861
/note="AT-rich; low complexity"
/rpt_type=tandem
/rpt_unit=ca
repeat_region 8334..8354
/note="AT-rich; low complexity"
repeat_region 9950..9970
/note="AT-rich; low complexity"
repeat_region 10777..10802
/note="AT-rich; low complexity"
join(<10872..11080,11797..>11973)
/gene="unkempt"
/mrna
/codon_start=1
/product="unkempt"
/protein_id="AAL73045.1"
/db_xref="GI:18463964"
/translation="YLKEFRTEDCPFLQHKCTQHRPFCFNMHFLNQRPRIRRD
GTENYSPDVCTKDETTGTCPPDCCPYLHFTTGTDEKRYHLRYKKTGTCTCHETDAR
GHVKNGLHCAFAHPDRLRPYDI"
17726..17763
/note="A-rich; low complexity"
repeat_region 19346..19371
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=c
20068..20121
/note="AT-rich; low complexity"
complement(<21113..>22747)
/gene="IGFALS"
/product="IGFALS"
/complement(<21113..>22747)
/gene="IGFALS"
/complement(21113..>22747)
/gene="IGFALS"
/codon_start=1
/product="IGFALS"
/protein_id="AAL73046.1"
/db_xref="GI:18463965"
/translation="ACARGCTCOHDYSLFLANYCSAONLTQVPPDMPSPSTSLMDV
NLFASLPASFKDLVHEFLNLSGOLVTLDPALGSLAHILHERRIRVLDGAI
FONMPKIASLSTNNOLSRIDRLPAGLSHMTLNGRSIVILPPTVRODQIGKEL
ILVGNRLAYIQPOLFOFNVFLKELDSGNLTKYIKANVYKILKLOKILANQIMTV
VPRFVGMKSRLMDLTNNRLTSLHEDTFEGLTLHLVLRVSNSTIGIAPRFRIQY
LEELRLSHNRIRVIGEKVEBELGLELEHNRIGEAIVGSGFTGLSHAVTNLSGSC
FHTLPDQVFGLSKLSHLDRGLTRVTQAFSGLSLRHLFTLGNNTSVYEROSFV
DIVGLIGLDFNKLEVLITGOTSGKLNLEYLLSNDCRHFLLONGSRLLPRLRID
LRANELTSIVPEAEGLKLLSGNRMRCCGALPLRNVSLRNALVVPQVETHAEGE
EPVTITIVNNITCHSPRLTGODLRDVESVQSC"
```

```
repeat_region 24332..24356
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=gga
repeat_region 24525..24554
/note="GC-rich; low complexity"
repeat_region 24862..24895
/note="AT-rich; low complexity"
repeat_region 26315..26340
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=ca
repeat_region 26915..26951
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=cgg
repeat_region 27240..27272
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=tcg
repeat_region 27487..27507
/note="AT-rich; low complexity"
repeat_region 28056..28080
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=tcg
repeat_region 30084..30122
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=cag
repeat_region 30217..30260
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=tcg
repeat_region 31181..31214
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=gga
repeat_region 33149..33186
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=tcg
repeat_region 33186..33241
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=gga
repeat_region 33298..33379
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=tcg
repeat_region 33966..34015
/note="T-rich; low complexity"
repeat_region 35457..35500
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=ca
repeat_region 35993..35993
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=tcg
repeat_region 37605..37637
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=tcg
repeat_region 37782..37804
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=ca
repeat_region 40208..40383
```

```
/note="simple repeat"
/rpl_type=tandem
/rpl_unit=tcca
40386..40562
/note="simple repeat"
/rpl_type=tandem
/rpl_unit=tcca
41620..41645
/note="simple repeat"
/rpl_type=tandem
/rpl_unit=tcca
41968..41994
/note="simple repeat"
/rpl_type=tandem
/rpl_unit=tg9
44729..44773
/note="simple repeat"
/rpl_type=tandem
```

```
alignment_scores:
  Quality: 4734.00      Length: 5488
  Ratio: 2.649          Gaps: 82
  Percent Similarity: 32.562  Percent Identity: 25.765
```

alignment\_block:  
US-09-736-968a-2 x AY016023/rev ..

Align seg 1/1 to reverse of: AY016023 from: 1 to: 167215

```
45 ValProLeuThrGluValValGluProLeuAspPheGluAspValLeuLeu 61
   ::::::::::::::::::::::::::::::::::::::::::::
134857 TTGCCCCGTGACAGAGGTGGAGCCGGATTGAGAGATTATGTGAG 134808
   ::::::::::::::::::::::::::::::::::::::::::::
61 uSerArgProProAspAlaGluProGlyProLeuArgAspLeuValGluP 78
   ::::::::::::::::::::::::::::::::::::::::::::
134807 CACCCATGCTCTCGGGGTGAGCCTGCTCCCTCAGAGAGCTTGTGAGT 134758
   ::::::::::::::::::::::::::::::::::::::::::::
78 heProAlaAspPLeuGluLeuLeuGluProArgGluCysArgThr 94
   ::::::::::::::::::::::::::::::::::::::::::::
134757 TCCCCCGACAGACGTGAGCTCTCCACCTGACAAAGAGTCCACACAC 134708
   ::::::::::::::::::::::::::::::::::::::::::::
95 ThrGluProGlyIleProLysAspGluLys..... 104
   ::::::::::::::::::::::::::::::::::::::::::::
134707 CTGGAGGCCCTCCCTGCTGATGAAGAGAGATTGTCCTCTTGCC 134658
   ::::::::::::::::::::::::::::::::::::::::::::
104 ..... 104
134657 AACCAATGCTCACACACTCTGCTTTCCTTTTAAACAACACACA 134608
   ::::::::::::::::::::::::::::::::::::::::::::
104 ..... 104
134607 CACACACACACATACACAGATGATGATTACATATTGAGTGAGAG 134558
   ::::::::::::::::::::::::::::::::::::::::::::
104 ..... 104
134557 TGAACATGACAGGACMAAGGCTTTTGAAGTGGCTGAATAACAAC 134508
   ::::::::::::::::::::::::::::::::::::::::::::
104 ..... 104
134507 ACACTTCTGATTAAAGACATTTTCTGCTCAATTATTCATGTTGTC 134458
   ::::::::::::::::::::::::::::::::::::::::::::
105 ..... 105
134457 TGTGTAGCTGACTGATCCAGAGTGGAGACACCTGTGCTAGTCTAC 134408
   ::::::::::::::::::::::::::::::::::::::::::::
118 GluAspTyrValIleValHisArg..... 125
   ::::::::::::::::::::::::::::::::::::::::::::
134407 GATGACTGGCTCATTCATTAAGAAAGTAAGTGTACATGTGAGGAGA 134358
   ::::::::::::::::::::::::::::::::::::::::::::
125 ..... 125
134357 CCGTTAATGTGTTCTCTGACAGACGCCCTTTTATCTGTGATTG 134308
```

```
126 ..... ArgTyrGlnTyrLeuSer 131
   ::::::::::::::::::::::::::::::
134307 CTTCGTTATTAACTCGCTCTCTTTTATGCTAGATCCAGCATACACG 134258
   ::::::::::::::::::::::::::::::
132 AlaAlaTyrSerProValThrThrAspThrGluArgGluArgLysGln 148
   ::::::::::::::::::::::::::::::
134257 ACCACTTTACAGCCTCAGAGCTCAGAGACAGAGAGAGAGGAGGAGG 134208
   ::::::::::::::::::::::::::::::
148 yLeuProArgGluValPheGluGlnAspAlaSerGlyAspGluArgSer 165
   ::::::::::::::::::::::::::::::
134207 GCTGCTCAAAACAGACCTTTGAGCTGATGATGGCGGCCCATGAGCAC 134158
   ::::::::::::::::::::::::::::::
165 LyrProGluAspSerAsn..... 170
   ::::::::::::::::::::::::::::::
134157 ACGACACAGTAGCTAATGTCATGTCATGATCGATCCGAGTCCATC 134108
   ::::::::::::::::::::::::::::::
170 ..... 170
134107 AGTTGAGATTATTATGACCACACCGTATTAAATATGATATGATGTC 134058
   ::::::::::::::::::::::::::::::
170 ..... 170
134057 GAGAAAGACATTACAGAGTTTTTTTCTTTTATATATGATAGCTGG 134008
   ::::::::::::::::::::::::::::::
170 ..... 170
134007 ACAAAAGACCAACCAACACTGAAAAAGCTGTCTCTGATGGAAAGAT 133958
   ::::::::::::::::::::::::::::::
170 ..... 170
133957 AGAAAAAGACAGATACACAGAAATGATATACCTGCTTATTCACCA 133908
   ::::::::::::::::::::::::::::::
170 ..... 170
133907 CATTCATATTATTAATTAATTAATTAATTAATTAATTAATTAAT 133858
   ::::::::::::::::::::::::::::::
170 ..... 170
133857 CATATTACTCTATATAAACAATCTCTGCAAGACCTTAAGGGGATG 133808
   ::::::::::::::::::::::::::::::
170 ..... 170
133807 TCCTGGAAGAAATTAACACATGATCTCATTCCTTTGCTCACCTCC 133758
   ::::::::::::::::::::::::::::::
171 ..... 171
133757 TTACGATCCACAGATGACGCAAGCGCGGTCACTCGAAGGAA 133708
   ::::::::::::::::::::::::::::::
182 hrProArgSerSerGlyAlaSerSerIlePheAspLeuArgAsnLeuAla 198
   ::::::::::::::::::::::::::::::
133707 CTCTCTGAGGAGAGCTGGGCTTCACGATCTTTGACCTGAAGAACTCC 133658
   ::::::::::::::::::::::::::::::
199 AlaAspSerLeuLeuProSerLeuLeuGluArgAlaAlaProGluAsp 215
   ::::::::::::::::::::::::::::::
133657 CCAAGATGCTCTCTCTGCGAGCTCTTACAGCAGCAGGCTCTGGAG 133608
   ::::::::::::::::::::::::::::::
215 LAspArgArgAsnGluThrLeuArgArgGlnHisArgProAlaLeuL 232
   ::::::::::::::::::::::::::::::
133607 GGACCAACGCAATACCGAGAGCGCCGACAGGGCCACAGTGCACCT 133558
   ::::::::::::::::::::::::::::::
232 euThrLeuThrProAlaProAspGlu..... 240
   ::::::::::::::::::::::::::::::
133557 TGGGCTTTACCAACCCCTGATGAGGTTTCTCAAAACATTTTGTG 133508
   ::::::::::::::::::::::::::::::
240 ..... 240
133507 TTGAATAATATGATTTTTTTTATATAACATGACATCAGAAATCTA 133458
   ::::::::::::::::::::::::::::::
240 ..... 240
133457 AATATTCTCAGCAAAACACACAAATTAATCAAAATTAATTAAT 133408
   ::::::::::::::::::::::::::::::
240 ..... 240
```

133407 AGATTGTCACTGGAATATGCCGGTCTGACAGATGCAATGTGATTTA 133358  
 240 ..... 240  
 133357 AGCCTTGTTGTTATTTTGAATAGCTTACATTTGACCTATGACATTTATG 133308  
 240 ..... 240  
 133307 CCACTGTTAAAAATGCATGGCCTTGAATGTTGTGATGTAGTGG 133258  
 240 ..... 240  
 133257 TATTGTTACGCTCTGGGTAAAGTAATGTCCTTGTCTCAAACTGCACC 133208  
 240 ..... 240  
 133207 CATTACAGAAGATAGTTTCGATTTGTCCTTGTCTTATACAGCTGT 133158  
 240 ..... 240  
 133157 GTTGACACATGTCACTTTCTGTGGTCCAGCTTGCATGACATGCTGT 133108  
 241 ..... AspGluAlaValGlu 245  
 133107 AAATATTTGTTCTGCTGTGTGCCCGCGTGACAGAGCGGGCTGTGAA 133058  
 246 ArgCysSerIleProGluProIleArgGluHisPheGluGlnArgIleLe 262  
 133057 AGATGCTCCAGCTGTTGAAAGTGCCCAAGAGCAGCTGTGCCAGAGATTAT 133008  
 262 uValIysCysLeuSerLeu..... 268  
 133007 GGTAAGTGTCTGTCTGAAAGTAGAGTAATAAATATTTAGATTATGT 132958  
 268 ..... 268  
 132957 CATTTATATGTCAAGCCCCCATTTTGGTAACATTTCTTCATACGCTTT 132908  
 269 Lys..PheGluIleGluIleGluProIlePheGluIleLeuAlaLeuTyr 284  
 132907 TCAAGTTTGAATATGAAATGAGCCGATATTTGGAATCCTTGCCTGTAT 132858  
 285 AspValArgGluIysIysLys..... 291  
 132857 GACATCAAGGAAAAAGAAAGTATTTTTCCTCATTTAGATGACTTGC 132808  
 291 ..... 291  
 132807 ACTTACGTAATGCTGAGCTCATGCTCCGCTAATTTCCCTTGTCTT 132758  
 291 ..... 291  
 132757 AAAAGGCTTAATAGCTCTACAGTTAGATCACCGAGTCAAAATGTAAT 132708  
 291 ..... 291  
 132707 AAATCCCTGACTTGAATTTCCATGATGCTCATCTCTTCCCTCTCCG 132658  
 292 ..... IleserGluAsnPheTyrPheAspLeuAsn 302  
 132657 CTTATCTTCCGCAACAGATCTCAGAGAAATTTCTACTTCGACCTGACT 132608  
 302 eAspSerMetIysGlyLeuLeuAlaHisGlyThrHisProAlaIle 318  
 132607 CAAATCAGATGAAGGGCTTACTTAACACACACAGCGCTCACACAGCCATT 132558  
 319 SerThrLeuAlaIleSerAlaIlePheSerValThrTyrProSerProAs 335  
 132557 TCTACCTGCGCCGCTCAGCTATTTCTCCATCAGCTACCCCTGCGCGA 132508  
 335 pIlePheLeuValIle.Lys..... 341  
 132507 TATCTTCTTGTCATTAAGTACTGAAGCTGCCACTTCACCTGATACATC 132458  
 341 ..... 341  
 132457 TGCTTAATGTAATGCCATTAATAATAGATCAGGTGTGTTGTGCAGACATGA 132408  
 341 ..... 341  
 132407 TCTAGTGTGCTAGATATACAAATGCAAAATTAATGTTTGGCGCTAGTCT 132358  
 342 ..... LeuGluLysValLeuGlnGlnGlyAsp 350  
 132357 TAACATGTTTGCATGCAATTTCACTGAAAAGGTCTCTCAACAAGAGAT 132308  
 351 IleSerGluCysCysGluProTyrMetValLeuLysGluValAspPheAl 367  
 132307 ATCAGTAGATGCTGTGAGCCCTTACATGTGATGAAAGAGTCTGATTCCT 132258  
 367 a.Lys..... 368  
 132257 AAGGTAAATGCAAAAAAATGTTACTGCTGTGATGTCACATAACAGCCACC 132208  
 368 ..... 368  
 132207 ATGTAAGTGCATGCTTGAAGTCAGAAACATTAATTCATTTCTGCTCCAAAA 132158  
 368 ..... 368  
 132157 GGTTCATTTTGTCTGTGTGGAACCTTTAATCTTCTGACAGCTACTCT 132108  
 368 ..... 368  
 132107 GAGGAGAGAGACAGTCACACATTTGCTGTAAATAAAGAGACAAAGTC 132058  
 368 ..... 368  
 132057 TGTGTAATAAAGAGTAATTAATTAACCACTGTCTGTGGGTGTGCGA 132008  
 368 ..... 368  
 132007 TGTTAATGTTTATACAGCTTTTGTGATTTCAAGTTCAACAATACAGT 131958  
 369 ..... AsnIysGluLysLeuGluLysLeuArgLeuAlaGlu 381  
 131957 TCCTGTTTCAGCAAAAGAGAGCTGGAGAGCTGCTCTCAGGACAGAG 131908  
 382 GluPheCysThrArgLeuGluIleValIleSerSerAlaGlyGlnLeuAspArg 398  
 131907 CAGTATGCAACCGCTTGGACGCTTGGACATGCTTTTGGCTGTGACAGC 131858  
 398 aValHisLeuAlaAsnIleValSerSerAlaGlyGlnLeuAspArg 413  
 131857 CATCCACCTGGTTACATTTGACAGCATATGGGACAGTCTTATGCGTACG 131808  
 414 ..... AspSerAspSer..... 417  
 131807 ACCCGACTTGTGATTCGCGTACCGGATTCAGCCACCTGTAAATGTCT 131758  
 418 ..... GluGlyLysArgPro..... 423  
 131757 AATGCCCTTTTACTGTCATGCTGAGTGAATATCTCCCTATATATAGCT 131708  
 424 ..... AlaTyrPheAspArgArgArgGlyP 433  
 131707 AACACAGAAAAAAGGCCATGGGACATGGAAATGAGACGAAGAAAGGGGG 131659  
 433 roGlnAspArgAlaSerSerGlyAspAspAlaCysSerPheSerGlyPhe 449  
 131658 ..TTTGACCGGATGATGTTTGCAGACGACATGTCAATTTTGCACACCTTC 131611  
 450 ArgProAlaThrLeuThrValThrAsnPhePheLysGln..... 462  
 131610 GGTCCAGCAACATTTGACCGGTACCAACTTCTTCAAGCAGAGTGTGACTGTCT 131561

132507 TATCTTCTTGTCATTAAGTACTGAAGCTGCCACTTCACCTGATACATC 132458  
 341 ..... 341  
 132457 TGCTTAATGTAATGCCATTAATAATAGATCAGGTGTGTTGTGCAGACATGA 132408  
 341 ..... 341  
 132407 TCTAGTGTGCTAGATATACAAATGCAAAATTAATGTTTGGCGCTAGTCT 132358  
 342 ..... LeuGluLysValLeuGlnGlnGlyAsp 350  
 132357 TAACATGTTTGCATGCAATTTCACTGAAAAGGTCTCTCAACAAGAGAT 132308  
 351 IleSerGluCysCysGluProTyrMetValLeuLysGluValAspPheAl 367  
 132307 ATCAGTAGATGCTGTGAGCCCTTACATGTGATGAAAGAGTCTGATTCCT 132258  
 367 a.Lys..... 368  
 132257 AAGGTAAATGCAAAAAAATGTTACTGCTGTGATGTCACATAACAGCCACC 132208  
 368 ..... 368  
 132207 ATGTAAGTGCATGCTTGAAGTCAGAAACATTAATTCATTTCTGCTCCAAAA 132158  
 368 ..... 368  
 132157 GGTTCATTTTGTCTGTGTGGAACCTTTAATCTTCTGACAGCTACTCT 132108  
 368 ..... 368  
 132107 GAGGAGAGAGACAGTCACACATTTGCTGTAAATAAAGAGACAAAGTC 132058  
 368 ..... 368  
 132057 TGTGTAATAAAGAGTAATTAATTAACCACTGTCTGTGGGTGTGCGA 132008  
 368 ..... 368  
 132007 TGTTAATGTTTATACAGCTTTTGTGATTTCAAGTTCAACAATACAGT 131958  
 369 ..... AsnIysGluLysLeuGluLysLeuArgLeuAlaGlu 381  
 131957 TCCTGTTTCAGCAAAAGAGAGCTGGAGAGCTGCTCTCAGGACAGAG 131908  
 382 GluPheCysThrArgLeuGluIleValIleSerSerAlaGlyGlnLeuAspArg 398  
 131907 CAGTATGCAACCGCTTGGACGCTTGGACATGCTTTTGGCTGTGACAGC 131858  
 398 aValHisLeuAlaAsnIleValSerSerAlaGlyGlnLeuAspArg 413  
 131857 CATCCACCTGGTTACATTTGACAGCATATGGGACAGTCTTATGCGTACG 131808  
 414 ..... AspSerAspSer..... 417  
 131807 ACCCGACTTGTGATTCGCGTACCGGATTCAGCCACCTGTAAATGTCT 131758  
 418 ..... GluGlyLysArgPro..... 423  
 131757 AATGCCCTTTTACTGTCATGCTGAGTGAATATCTCCCTATATATAGCT 131708  
 424 ..... AlaTyrPheAspArgArgArgGlyP 433  
 131707 AACACAGAAAAAAGGCCATGGGACATGGAAATGAGACGAAGAAAGGGGG 131659  
 433 roGlnAspArgAlaSerSerGlyAspAspAlaCysSerPheSerGlyPhe 449  
 131658 ..TTTGACCGGATGATGTTTGCAGACGACATGTCAATTTTGCACACCTTC 131611  
 450 ArgProAlaThrLeuThrValThrAsnPhePheLysGln..... 462  
 131610 GGTCCAGCAACATTTGACCGGTACCAACTTCTTCAAGCAGAGTGTGACTGTCT 131561



462 ..... 462  
131560 TGTCAATAGTAAAGCTTGAGTGTGGCTATATTTGATTTGTTGCTATC 131511  
462 ..... 462  
131510 ATCAGCTATACATACAAACAAATGATGATTTATCTTAAGACACCACCTA 131461  
462 ..... 462  
131460 GACTGTGATCAAGTAAATGTATGCAACTCAAGTTATAGGAAAAAAGT 131411  
463 ..... Glu..AlaGluArgLeuSerAspGluAs 471  
131410 TAAATTTCTTGTGTTTCATGAAGGAGGGGACCGAGCTAAGTGAAGAGGA 131361  
471 pleuPheLysPheLeuAlaAspMetArgArgProSerSerLeuLeuArg 488  
131360 CCTTTATAGTTTCTGGCAGATATGGCAGACCGCTCTCTGTTCTGGCAGC 131311  
488 rgleuArgProValThr ..... 493  
131310 GACTGAGGCTGTGCACAGGTAAAGCAATTACTTGTACAAAGATA 131261  
493 ..... 493  
131260 ATTGTTCTAAGATGTATGTGTGTTGTCTCTCTAACTTAAGATGC 131211  
493 ..... 493  
131210 ACGTTAAAGTGSTTAATCCATTTTCTTCTTTGTCTGTGATTTAT 131161  
494 ..... AlaGluLeuLysIleAspIleSerProAlaProGluAsn 506  
131160 TTGACACTGCAGCCAGTTGAAGATTGATTTCTCCAGTCCGGAGCTGG 131111  
507 ProHisPheCysLeuSerProGluLeuHisIleLysProTyrProAs 523  
131110 CCTCATTTACTGTCTGTACACAGAACTGCTTCATGAATAAACCTACCCACA 131061  
523 pProArgGlyArgProThrLysGluIleLeuGluPheProAlaArgGluVal 540  
131060 CCCAGGGGTGGCCCAACCAAGAGTGGTGGAGTCCCTGTCGTTATG 131011  
540 alTyrAlaProHisThrSerTyr ..... 547  
131010 TGTACACGCCACATACCTCTATAGTGAAGGCCCCCAACTCACAGAA 130961  
547 ..... 547  
130960 TCATTTATTTCTTCTCTCATATCAATTAATTTATTTGTCGCTTAT 130911  
547 ..... 547  
130910 TTGCTTATTAATCTGATAAATATTAAGAGATAAAAACACAGTAATACCT 130861  
547 ..... 547  
130860 ATTGAAAAACAATATTTATATACATTTCTAAATAGAAATACATTTCTCAG 130811  
547 ..... 547  
130810 TAAATTTGCATGTTTACTCAAGTTTAAATATGTAACAATTTGTTCAAGTAA 130761  
547 ..... 547  
130760 TAAAGCTAAGTTCACAAAGTACAGACAAACTCCCATCATACAGGTCAAA 130711  
547 ..... 547  
130710 CGATTAAGAACTTTGTAGTAAAGTCTATAGTCAAGTTCAGGCTTACGTACA 130661

547 ..... 547  
130660 TGTACGGAGAGAAAGAGATGCAGTATGTTTGCAGTAAAAAAAACCTG 130611  
547 ..... 547  
130610 TGGCGAATCTTAATGGCCAATGCTTGCCCTGCTCTTGCAGATGCAAA 130561  
547 ..... 547  
130560 GAGAAAAACAGATACAGCATTTAGAGGATATGAATAAACACGGGGCGTT 130511  
547 ..... 547  
130510 TTAACAATAATCTTACCGTTTCTGTAGTTAGTGAAGCTTTCACCTG 130461  
547 ..... 547  
130460 ACTCTTGATGCTCTGCTTTTGAACAGTAATCAGTTCTTGTAGTCTCTCA 130411  
547 ..... 547  
130410 AAAGTTGAGTGTCTCTGCTCTAGTCTAGAGCGGGGAGAACAGTTTC 130361  
547 ..... 547  
130360 CTAAATCATTTTTTGTGTGTATGCTGTAATTCCTTGTAGTAAATTA 130311  
547 ..... 547  
130310 CAGAGGAATTTCTTTACTGCAGCTCTTTGATGCTGCTGGTGGGTACATA 130261  
547 ..... 547  
130260 CTTATTGAGATGTGTCACAGCTCTTTACATTTGCTCAGTCCACTT 130211  
547 ..... 547  
130210 CTTCTGCTCCAGTTTGGCAGCCTGTCTTTCGTCACTCGGTCAGACC 130161  
547 ..... 547  
130160 TGTTCCTGGCCATCTGGGTCTGTTGAAGAAGCATCTGTCCATGTTATTC 130111  
547 ..... 547  
130110 CTCTACAGCAGCAATCTCACGGCAGTGGCAGCGGCTGACCCCGTAT 130061  
547 ..... 547  
130060 AACATGTGGCTTTCTAATGACACACTCTCAGATGAAGGGTCTCTATAT 130011  
547 ..... 547  
130010 CTGCAACATCATGAAAAACAAATGTACAGCAAAATACCAAAAGTCTTA 129961  
547 ..... 547  
129960 GAAAAACAATCTTATCCCAAGAAAGACAGTGTATTACAGATTTTGAG 129911  
547 ..... 547  
129910 AATGACCCCTGTGTTTCTGGCCCTCTGCATGTTTCGGTCTCAGATT 129861  
547 ..... 547  
129860 GATTACAGTGTGTCTCTCTCTCTAGAGACAGCCCTTCAAGTAG 129811  
547 ..... 547  
129810 CCAGCAGACATGTTGATGACATATTTCTGCCATCTTACCTGTAGCAA 129761  
547 ..... 547

129760 CATTATACTGCTTAGCCGTTTTCATCTCTGACGCTTCTCAGCAT 129711  
 547 .....  
 129710 GTCTCCCAAGCCTTGAGATATCCTTTCATGGTTCTCAAGTCTGGCTT 129661  
 547 .....  
 129660 ATTCTGTTATCTTGGCTCGGTGTTTCAGAAATGATTTAGGGAGTC 129611  
 547 .....  
 129610 GCCGAAGTATGAATCCAAAGCTGAGCAGTTGACCTGCTCTGGTG 129561  
 547 .....  
 129560 AAACATTGTTTTCTGCTGCTCTCTCAGTGGACCTGCAGGGTGTCT 129511  
 547 .....  
 129510 CTTAAACCCCAAGCCAAACCATCAGACAGACTCCAGATCAGATTTCAT 129461  
 547 .....  
 129460 AGTATATTTCCCTCACTGATGACGCTGTTGACAGTTCTGTGTTTC 129411  
 547 .....  
 129410 TTCTCAGATGTTAGTCAAGAAAGTTAGTAAATTCGTAATAGAGT 129361  
 547 .....  
 129360 CCCATCTGTTTAAGATCCAGTGTGTGTGGCTGTGTGAATTTGGAGG 129311  
 547 .....  
 129310 GTAGTCAGTGCACTTATATTTGATGTGATATGATAGTAAATATTA 129261  
 547 .....  
 129260 GCTCAGCTGTGATGTTCAATCAGTGCACCTTACTCTTTGACCTTGCCTTA 129211  
 547 .....  
 129210 ATGTTTAAACCCCTGAGATCTGATCTTGACATTTAGGTCACAAACATG 129161  
 547 .....  
 129160 TGACACATTACTGTATATTTTGTGTTTGACACAGATCTTGGTGTGTTT 129111  
 547 .....  
 129110 TCATGCCACACTGCTGTGTAATGTGTAAAAAGATATTTCATAGTGCC 129061  
 547 .....  
 129060 AAACATCTTACTCAATPACAAAGAAATTACATCAGCTGTCCAAAGTAC 129011  
 547 .....  
 129010 CTCGATGTTGACTTTAGTTTGTGAAGTTACGGTGCATTTGACTCCACA 128961  
 547 .....  
 128960 ACATCCAAATGCTGTGCACACTCTGTACATAGATTTTGCACATTCAT 128911  
 547 .....  
 128910 GAATATCTTTCATTTACTCTTCTAAGATTGTTAGTATGACAGTTGTA 128861  
 547 ..... 547

128860 GATGCTTTGACAAACAGAAATGCCAACTTTCATTCATATTTAAATTT 128811  
 547 .....  
 128810 TGGGAACAAGATCTTCCCTTCAAAATCAGCATTAACAAAAATATATGTA 128761  
 548 ..... Arg, AsnLeuLeuTyrValTyrProHisSerLeuAs 559  
 128760 AAACCATGTCATTTGCAGGAACTGTCTATGTATGTTACCCAAAAGCTTGA 128711  
 559 nPheSerSerArgGlnGlnSerValArgAsnLeuAlaValArgValGlnT 576  
 128710 CTTCAGTAGTGTCTCAGGGCTCAGTGAGGAACATCGCTGTCAAAATCCAGT 128661  
 576 yrmEtrhGlyGluAspProSerGlnAlaLeu..... 586  
 128660 ACATGCGAGAGAGAGACCCAGCTCAAGCTTTGCAGTGTAGCAGAAATCAA 128611  
 586 ..... 586  
 128610 GGACTACCAAGTGAATAAATAACATGTTCTAAGAAAAACAGACCGAAT 128561  
 586 ..... 586  
 128560 CTGTATTTCTGAGAAATTTACCTGGCTTTAGTAAATCTGCAAGTAAGTAAAA 128511  
 587 ..... Pro, ValIleP 590  
 128510 AAACATCCCTAAATGATTTTGGTATTTTATTTTACCTAGGTCAATTT 128461  
 590 heGlyLysSerSerCysSerGlnPheThrArgGlnAlaPheThrProVal 606  
 128460 TTGGAGTCAAGTTGTGTGATGTTCAAAATGAGGATACAGTCCATC 128411  
 607 ValTyrHisAsn..... 610  
 128410 ATCTACCAACAACAAGTAAAGCTTTCACAGACTTAACTGCTGTAACTCC 128361  
 610 ..... 610  
 128360 TTTGTGTTTCCCTAAATCTGTGTGTTTAACTCTTTTGTGTTAAAAA 128311  
 611 ..Lys..SerProGlnPheTyrGlnGluPheLysLeuHisLeuProAlaC 626  
 128310 AAAAAGTCCCTGAGTGTATGAGGAGATCAACATGAAATTTCTCTGCCA 128261  
 626 ysValThrGluAsnHisLeuLeuPheThrPheTyrHisValSerCys 642  
 128260 ATCTGACTGACATCACCACTCCGTGTACCTTCTACACATCAGCTGC 128211  
 643 GlnProArgProGlyThrAlaLeuGlnThrProValGlyPhe..... 656  
 128210 CAGCCCAAAACAACACATCTCTAGAGACTCTGTGGCTACACTGTGAG 128161  
 656 ..... 656  
 128160 TTCACTTAGTCAGTCTCATTTTATCTCCACCCACAGCCCAAGTTT 128111  
 656 ..... 656  
 128110 GATAATTTCTGTTGTGCATGCATCTTTAAAGATCTCGTGTCTTCTT 128061  
 657 .....Thr..TyrIleProLeuLeuGlnHisGlyArgLeuArgThrGly 670  
 128060 TCGACTGTAGTGAATCCCTGTGATGCGACGCGCCGACTGAGAAGCAGCC 128011  
 671 PropheCysLeuProValSerValAspGlnProProposerTyrSerVa 687  
 128010 TCCTTCAATTTACCTGTCTCAGTGAAGAGCCCCACGAGCTACTCTGT 127961  
 687 IleuThrProAspVal, Ala..... 693  
 127960 TCTCACCCCTGATGTAAGCAACGACTGGGATATCAGTTTGCCTGAAT 127911

```
693 ..... 693
127910 GACCCTTGGCGTTTGTACTGGCTGTTAATGAGATGGATTTGCGTGGAT 127861
694 ..... LeuProGlyMetArgTyrPValAspGlyHisLysGlyValP 707
127860 GTAGGTCACGCTTCCAGGATGAGTGGTGGATTAATCAACAAAGAGATGT 127811
707 heserValGluLeuThrAlaValSerSerValHis..ProGln..... 720
|||||.....
127810 TTAAGTTGAAGTACACAGACGCTTCTCGGTTCATCTACTCAGGTACACGAC 127761
720 ..... 720
127760 GCTAAATATTGTGCTTTCATTCATCATCATCTGACACGATCATTTCAAG 127711
720 ..... 720
127710 CATCCATTGATCTTAATGTTAAGTAATTAATTAATAGTCTGTAAC 127661
721 .... AspProTyrLeuAspLysPhePheThrLeuValHisValLeuGluG 736
|||||.....
127660 GTAGGACCCCTCTCGGATTAAGTCTTCACGCTGCTGTATGCTCTGGAAG 127611
736 LuGlyAlaPheProPheArgLeuLysAspThrValLeuSerGluGlyAsn 752
|| .....
127610 AGTACTCTCCCTCCGCTCCGCTGAGGATGTGATCATTTACTGATCAAC 127561
753 ValGluGlnGluLeuArgAlaSerLeuAlaLeuArgLeuAlaSerP 769
|||||.....
127560 ATGGAGGCAAGTTAAAGACACAGCATGTGGCCTGACGAGGGGCGCTGTT 127511
769 ogLupProLeuValAlaPheSerHisValLeuAspLysLeuValArgL 786
|||.....
127510 GGATACCTGTGTGACGCTTCTCCACACGCTGCTCAACAGATCATTTCAAC 127461
786 euValIleArgProPheIleIleSerGlyGlnIle..Val..... 798
|||||.....
127460 TCATCGTTTTCACCGGTCATCGAGGCCAGCATGTGTAATGAGCTGTG 127411
798 ..... 798
127410 AAGCACCGTCGTGCTGTACTGCTACAAATTGTGATTTGGAGTGTGCTT 127361
798 ..... 798
127360 CTGTGAACAGGCGTGAAGGCTCATTTCTAATAGTGGGTTAATAGTTGAT 127311
798 ..... 798
127310 GGTAATTGATCAAAATTACTGTGATGAAATAATATGACATTTTGTCTCA 127261
798 ..... 798
127260 TACACATAAAACACTGACTTTCATGTCATGCTAAACATGTTGTTTAA 127211
799 ..... AsnLeuGlyArg 802
|||||.....
127210 AATAATCCTACATCTTGATTTCCAACTCCTGACAGTCAACCTTGGCAGG 127161
803 GlyAlaPheGluAlaMetAlaHisValAlaSerLeuValHisArgSerLe 819
|||||.....
127160 ACTGCCCTTTGAACTATGGCTTACTGTGTAACACACACACAAACCT 127111
819 uGluAlaAlaGlnAspAlaArgGlyHisCysProGlnLeuAlaAlaTyrV 836
|||||.....
127110 GGAGGGAACAGACCATATCATGTGTGCAACAACCTGTGGCTTC..TACA 127062
836 alaHisTyrAlaPheArgLeuProGlyThrGluProSerLeuProAsp... 851
|||||.....
127061 TCGAATATATGCTTTGCGCTGCCACATGCTGAACATCATGTGCCCAACA 127012
908 ..... 908
```

126111 TCCTGGCCCCGCCCCCTTCTTGTGTTTTCACATTATATCCTCC 126062  
908 ..... 908  
126061 AAATTAATTACTGTGTGTGTAATTAGCTTCCCTCAGCTCAGTACCCTTG 126012  
908 ..... 908  
126011 CCTGTGCCACTGTGTGTTTGTGTTGCTTCTCGTGCGTTAAGCAGG 125962  
908 ..... 908  
125961 CATTGACCGGCTCCACTCTGTGGTTAACTCTGCTATGCCCGGGGCTT 125912  
908 ..... 908  
125911 CCAGATCATTTGCTTGGCGGAAACCCCACTCCAGCTGTGATCGAAGCAG 125862  
908 ..... 908  
125861 GTGCCAATCCCTCTCTCCGCTCCTCTACCTACCTGCGTCCGCGCAGTT 125812  
908 ..... 908  
125811 CTAAACAACACTCTTGAGTCATTTGTTCTGCAATAGCTTCATCTGCT 125762  
908 ..... 908  
125761 TTTCACTGTGCCACTCCTCTTATCTGTGAATACTACTATCTGTGTCAC 125712  
908 ..... 908  
125711 CATTCATCCTCTCAGACACCATAAAGCCACATCTGTGTGAGTGCT 125662  
908 ..... 908  
125661 GCCTGTGTGAGATTTTCTACTAGCCTCAGCTCATGCTACAGAGTCTG 125612  
908 ..... 908  
125611 TGGTACCACCAATTTTCCATTTCATCCATTTTCACTACATTTTGTCTGTG 125562  
908 ..... 908  
125561 TGTTCACCTCAACATCACACATGTGACAGCAAGCGCAGCTGCAATC 125512  
908 ..... 908  
125511 GCATGTGTGCTTCTCTGGACAGCATGGCTTGTTCAGTTCGCCAAGG 125462  
908 ..... 908  
125461 CACATTGCCAAGAGGTAAAACTGTCTTCTACTATACAGAAATACAGAGT 125412  
908 ..... 908  
125411 GTTCAACTCAAGCGATGATTTTGTGCTCAGCATGATTCATTTTCA 125362  
908 ..... 908  
125361 CTGTATCTCATTCATCGTACATCACCTCATATGTAGTTAAATGCTGA 125312  
908 ..... 908  
125311 CACTGATCTAAACATTTTAACCAAGTATGTATTTATTCATTTATTC 125262  
908 ..... 908  
125261 CATTTTAAATTAATTAATTAATTAACACTTAATCTCAGACAGCTGT 125212  
908 ..... 908

125211 AATGTTTAGATGTTTCTGTTTTCATTTAAATCTAATTCATATTTT 125162  
908 ..... 908  
125161 AACATTATATTAATGACAAAATTTCTAGTTAAGCAGTGTGCAAAAC 125112  
908 ..... 908  
125111 TTTCTAGTGTGTTTGTTCCTGTAAGTGTATTCACGTTTATGTCAT 125062  
908 ..... 908  
125061 GGAATATTTAGATTTTGTAAATATATGACTCTCTGATTTGGACCGC 125012  
908 ..... 908  
125011 AATACCTTTAGAGTTGTGCAACACCTTTTGGAAATGCTCTACAGTA 124962  
908 ..... 908  
124961 AATGGTTAATTAGTAATCTGTATCTAAGGAACTTTGTAACATTTCA 124912  
908 ..... 908  
124911 CAGTAAGGTGGCTAAATTTATGGCTTCTAGGAATCATAGTAAAA 124862  
908 ..... 908  
124861 CAGCAAAACATTTAATGACGACGAGATTAATTAGTTAGCAGCTGTG 124812  
908 ..... 908  
124811 ATGTTTTCGAACGTCCGATATTGACCAAAAGTATGATATTAAC 124762  
909 ..... HisGluJluLeuAlaL 914  
124761 TGACTTATGTGTGCTGCTGCTGTCAGTTGCTTCATGAGAGTTGCTT 124712  
914 eugIntpValIValSerSerAlaValArgJluAlaLeuGlnHis 930  
124711 TGCAGTGGGTGTCAGCACGCCACGATGAGGAGCAGCGCTTCAGCAC 124662  
931 AlatrPhePhePheGlnLeuMetVal ..... 939  
124661 GCCGTGTTTCTTCCAGCTGATGATGAGCATGCTGTGTGTCTGTCA 124612  
939 ..... 939  
124611 AATCCATTAACACATTTTAAATTAATAGTCTGAAAGTCCCATTTCTC 124562  
940 ..... Lys..SerM 942  
124561 TTTCCTCTTTCTTTTACTCTTTTACTCTTCCCGACGACAAAGATTA 124512  
942 etAlaLeuHisLeuLeuGlnArgLeuAspThrProArgLysLeu 958  
124511 TGTCTCATCACTTATTTCTGACCTCAAAAGTGTATCCGAAACGCCAA 124462  
959 ArgPheProGlyArgPheLeuAspAspIleThrAlaLeuValGlySerVa 975  
124461 CGATTCCAGACCGCTTGTGGATGACATCGTCGCTGTGTGTGTCAT 124412  
975 IGLyLeuGluValIleThrArgValHisLys ..... 985  
124411 CAGTGCAGATATTTCTGTGTCATCCCAAGATCTTGAGATTGCGTACA 124362  
985 ..... 985  
124361 TTTCCTTCAAAAGCTGAGGTTTTTTTAGTGTAATAGCATTAAGTCAA 124312  
985 ..... 985  
124311 AACATAATTGATCGGTAAATATGTGAACATCTCTTAAACATCCTCT 124262

986	.....	Asp	967
124261	TGCCCTATTTTCACATGAACCTCCACAGCTTTTCTCCCTGTACAGAGAT	124212	
987	alglutenuaIagluHisIleuasnIasertleuAlaIapheluSerAsp	1003	
124211	TTGAACTTGTGGAGGGTTAAATAGCACTGTGGCTTCTTCTGTAATGAC	124162	
1004	LeuIeuSerIeuValAspArgGlypheValIphesertleuValAlaIaH	1020	
124161	CTGCTTTCCTCATGGACCGAGGCTTTGTATTGCAACCTCAATCCGCTGA	124112	
1020	styrIysgluValAlaIaThrArgIeuGluSerPro.....	1032	
124111	CTACAAACAGAGTGTGCTCT.....TCTCTCCGTTGTGCTTAT	124069	
1032	.....	1032	
124068	TATCAACACCTTTCCTTCACATATATTTGTCTCAGTGTGTGGCTT	124019	
1033	.....AspProAlaIaIeuLeu	1038	
124018	TAAACAGATTGGCCAAACAGCTTCACACGCGACAGAAATCCAGCTCCGAA	123969	
1038	uThrIeuArgMetGluIuIpheThrArgIleuIeuSerHisGluHisIyry	1055	
123968	TGCCCTGAGAGATGGACTTTATCCGATGTGTGTAGCCATGAGCAGCTACG	123919	
1055	alThrIeuasnIeuProCysCysProIeuSerProProAlaIaSerProSer	1071	
123918	TCATCTCTCAACCTGCGGTTCCTACCTCTCACCCCTCCAGCCTGCGCTCA	123869	
1072	ProSerValSerSerThrThrSerGlu.....	1080	
123868	CTTTCACACTCTCTCCACACGTCACAGGTGCTATTAATGCTGGAGGCT	123819	
1080	.....	1080	
123818	TAGTACAAATTCATTACATTGCTCACAATATCCAAATCTTTTCTTCA	123769	
1081	.....SerSerThrPheSerSerGI	1087	
123768	TTTTCACTTTCTTTATTTTACGTTCCTCAGAGTTCCAGACTTTTCAGCAT	123719	
1087	nalAProAspProIySValIThrSerMetPheGluIeuSerGlyProPhea	1104	
123718	GGTGAAGAGATCAAGAGAGTGGCCACAAATTTGAGACTCTCCGTCGGCTTC	123669	
1104	rgIuIngIuHisPheIeuAlaGlyIleuIeuIeuThrGluIeuAlaIeuAla	1120	
123668	GCCATCAACATTTCCCTCTGTGGCCGCGCTTATCAGAGCTCTGCTCATC	123619	
1121	LeuGluProGluAlaGluGly.....	1127	
123618	CTTGATTCCTGATGTGTGAAGGGTAGGAGTGAATCTTACTTTTTTTTTT	123569	
1127	.....	1127	
123568	ATGTCTGGCTATTTGGACCATGTCTAAATGATTACTGAAATAGCTGTATG	123519	
1127	.....	1127	
123518	CTTCAAGTCCAGTACATGAAACCTGAGTTGTGGTTTTTTCTACACACAG	123469	
1128	AlaPheIeuLeuHisIySValAlaIleSerAlaValHisSerIeuIeuC	1144	
123468	CATTTCCTTCCTCCATAAAAAAGCATTAAGTGGCTGTACACTCCCTGTGT	123419	
1144	ysGlyHisAspHisAspProArgIyAlaGluAlaIuIuValIySAlaArg	1160	
123418	GCAGGCATATGAGATGTTCTGCCCTCAAGAGACCTCAGAGTCAGGGCCAT	123369	

1161	ValAlaLeuLeuTyrLeuProLeuLeuSerIleAlaArgAspThrLeuP	1177
123368	GTTCCTAGCTTTAACTGGCCCCCATCTCCAAATGGTATGGAGACATGAA	123319
1177	oArgLeuHISAsPheAlaGluGlyProGlyGlnArgSer.....	1190
123318	TCAGCTTCATGACTTGTCT.....GGTACGTGTACACCCCCACCA	123278
1191	.....Arg.LeuAl	1193
123277	TTTTCAATACCTTGGCAAAATGTCACCTTGTGGCATTTAAAGATGTTAAAT	123228
1193	AspMetLeu.....	1196
123227	TTCAATCTGTGTGTATGTTTTGTGTAGACTCTCCCTCGGGGTC	123178
1197	..AspSerAspThrGluGlyGlnAspPhe....AlaGlyThrIle	1210
123177	GGGATGGCTTAGCCCAAGACAGATGGCAGCTGGATGGATGGCCAAACCATC	123128
1211	AsnProSerValAlaMetAlaIleAlaGlyProLeu.....	1223
123127	AGTCAGTCTGTCCATGGCAATGGCCGCTCCCTTTACACATGCCAA	123078
1223	.....	1223
123077	AGCCAACTCTTTGGCGCTACCCACAGCGGTAGTATATATGGCTGGGGTC	123028
1224	....Ala..ProGlySerArgAlaSerIleSerGlnGlyPro.....	1235
123027	AGGTTGTCCCTGGCGCATTTACAGAAACAGTCACAGATCCCACTGTGTG	122978
1235	.....	1235
122977	TGGAGACTGTGGCTCAGCTTGACAGAAACACACACTGGGCTTTCGTA	122928
1236	...ProThrAlaSerArgAlaGlyCysAla.....	1244
122927	ACTCCACTAAAAAAAACACAAAAATGTGCATTTCTTCCACATGCTCATC	122878
1244	.....	1244
122877	TTCCCTCAAGATAACTGCTGTGTGTGTGCTTACGTGGCTCTGTGGCCT	122828
1244	.....	1244
122827	CTGTATAACTGTCTTTTCATGTTTGACATTGAAAGGSCAACTTAGCGTGTG	122778
1245	.....LeuSerAlaG	1248
122777	TGTGTGTGTGTGGCAGCGCTGGCGCCAGTCCACCTCTGTGTGACAGA	122728
1248	userSerArgThrLeuLeuAlaCysAlaLeuThrPheAlaLeuLysAsnArg	1265
122277	GTTAGCGCGGACTCTGTGTGTGTCTCTGTGGGTGCTGAAGATACAG	122678
1265	IuProAlaLeuLeuGlnArgTyrPheAlaThrAspLeuThrLeuProGlnLeu	1281
122677	ATGCAAGCTCTCCGAGAGCGGTGGGTGTGATTTGTCAAGTCTGCACAAATC	122628
1282	GlyArgLeuLeuAspLeuLeuTyrLeuCysLeuAlaAlaPheGluTyrGly	1298
122627	AACGCGCTGTGATCTGCTGACGCTGTCACTCTCTTTTGAATATAA	122578
1298	.....	1298
122577	AGTAAAGTACTGTGTGATTTTGTGCTTTGTTTCACCTCTTGAGTCTGT	122528
1299	.....GlyLysLysAlaPheGluArg	1305
122527	TGACATGTTTAACTTTTGTATTTAATAGGGGAAAAAGGCTCTGTGAAGG	122478
1306	IleAsnSerLeuThrPheLysLysSerLeuAspMetLysAlaArgLeuG	1322

122477 AFTAAATACCTGACATTTTAAAAATCCCTGGACATGAAGCAGCGGTGGA 122428  
1332 uGUuAlAlleuGlyThrIleGlyAlaArgGlnGluMetValArg. 1338  
122427 GGAGCCACTACTGCGAACCATTGGGGGCTGCTCAGAGATGGTTGTGCTT 122378  
1338 ..... 1338  
122377 GCAGAGTAATATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122328  
1338 ..... 1338  
122327 TGT 122278  
1339 ..... Ser ArgGluArgSerProPheGlyAsnProGluAsnValArg 1352  
122277 TCTCAAAATCTTTTAGAGAGGAGTCCCTATAGCAGCTCAAGAAATGTGAGG 122228  
1353 TPATrgLysSerValThrHisTrpLysGluThrSerAspArgValAspLys 1369  
122227 TGGAGAAAGAACGTCACCTGCTGGCGCAAAATACAGACAGAGTCGACAA 122178  
1369 s..Thr..... 1370  
122177 GTATGTAA 122128  
1370 ..... 1370  
122127 AGATGTTTCATGTGCGTATAGCAACTAGTGTTCATCAGCGGTGCG 122078  
1370 ..... 1370  
122077 GATCACTGTGCTTTTAATTTTCCAAGTGTGAAATGCTGATGCTGAAGC 122028  
1371 ..... LysAspGluMetGlnHisGluAl 1378  
122027 TAAGCAATTAATTAATCTTCTTAGAGCTAAAGCTGAGATGGAGCGAGTGC 121978  
1378 aLeuValGluGlnLysLeuAlaThrGluAlaSerLeuValValLeuAspT 1395  
121977 CGTGTGTGAGGAACCTTATCAACCGAGGCTTCTTAATGCTGCTTGACA 121928  
1395 hrLeuGluIleLeuVal.Gln..... 1401  
121927 CGCTGGAGATTTGTAGTCAAGGTAGACAATGCTTCAAGCTGTGAAGTGC 121878  
1401 ..... 1401  
121877 GAAATTAATGAGAAATTAATTTCTGTTCTTCTGCTGTTTGATTTTTC 121828  
1402 ..ThrValMetLeuSerGluAlaArgGluSerValLeuGlyAlaValLeu 1417  
121827 AGACTGTGTGGCATGACAGCTGAAGGAAAGTGTGCTGGTGAAGTAATG 121778  
1418 LysValValLeuThrSerLeuGlySerAlaGlnSerAlaLeuPheLeuG 1434  
121777 AGAGTCTCTCTCCACAGCATGGCAGGAACCAAGAGTCCCTCTTCTGCA 121728  
1434 nHISgLyLeuAlaThrGlnArgAlaLeuVal..... 1444  
121727 GCACCTGCTTCAACACAAAGAGCTCTAGTTTTCAGGTATCAATTAAT 121678  
1444 ..... 1444  
121677 TTTTGTGTTAATTTGAGATTAATTTTCTTCTTAATCTTTTACATGTT 121628  
1444 ..... 1444  
121627 GCAGCATATTACATCATAGTTAGATGACATACATATAGATAGACA 121578  
1445 ..... Ser..Lys..... 1446  
||| |||

121577 TGTGTGACCACCTTGAAGTGAAGCAACTGAATATACATGTGTGT 121528  
1447 ..... PheProGluLeuPheGlnGluAspT 1456  
121527 TTAATATACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 121478  
1456 hrGluLeuCysAlaAspLeuCysLeuArgLeuLeuArgHisCysGlySer 1472  
121477 CAGAACTTGTGCGCAGCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 121428  
1473 ArgIleSerThrIleArgThrHisAlaSerAlaSerLeuTyLeuLeuMe 1489  
121427 AGCTGATGTTCTGTGCAGAAAGTCAAGCTGCTGCTGCTGCTGCTGCTGCT 121378  
1489 tArgGlnAsnPheGluIleGlyHis..... 1497  
121377 GACACGAAGACTTTGACATTTGAAATGTAGTAATATTAATCTTCAATGTCA 121328  
1497 ..... 1497  
121327 TGATGTAAATTAATTAATTTCAATTAATTAATTAATTAATTAATTAATTA 121278  
1497 ..... 1497  
121277 CTGACATTTACTTAAAGCTGAGACAAAGAAATTAATTAATCCTT 121228  
1498 ..... AsnPheAlaArgValLysMetGlnValTrpMetS 1509  
121227 TTAATTCCTTTTGACAGACTTGTCTGGTAAAGATGACAGTCAACATGT 121178  
1509 erLeuSerSerLeuValGlyThrThrGlnAsnPheSerGluGlnHisLeu 1525  
121177 CACTGTCAATCACTGTGTGGGAGACATCAAGAACTCAATGAGGACATCTT 121128  
1526 ArgArgSerLeuLysThrIleLeuThrTrpAlaGlnGluLysAspMetGlyLe 1542  
121127 CGACACTGCTTGAAGAACTATCTTGACATATGCTGTAAGATGATGTGAAC 121078  
1542 uArgAspSerThrPheAlaGluGln..Val..... 1551  
121077 AGGTGACACACCTTCCCGAGCAGGTATTAATGCGCAGTGAAGAT 121028  
1551 ..... 1551  
121027 CCATTTGTTGAAGATGATGAATGTTGAGATTTTAATTCATTCCTGTGT 120978  
1552 ..... GlnAspLeuMetPheAsnLeuHisMetIleLeuThr 1563  
120977 TGAACTGACAGTCCAGATTTGTGTTTCAATTTGCATGATGTTCTCAC 120928  
1564 AspThrValLysMetLysGlnHisGlnGluAspProGluMetLeuIleAs 1580  
120927 GATACGTGAATAATGAAGAACACACAGCAGATCCAGAGATGCTTATGA 120878  
1580 pleuMetLysArg..Ile..... 1585  
120877 CCGTGAATGTAAGATATGAGATGACATTAATCAAAATATAGTATCA 120828  
1586 ..... AlaArgLysTyArg 1591  
120827 GACTTAATATTTGCTTTTGTAAATACAGATTTGCAAGAGGCTATTCAA 120778  
1591 LysSerProAspLeuArgLeuThrTrpLeuGlnAsnMetAlaGlyLysHis 1607  
120777 ACTCTCCGACCTGCTGCTGACTTGTCTCCAGAACATGCGAAGAAACAC 120728  
1608 AlaGluLeuGlyAsnHisAlaGluAlaAlaGlnCysMetValHisAlaAl 1624  
120727 TCTGAGAGAGAACCATGCTGAAGCTGCCACCTGCTGCTGCTGCTGCTGCT 120678  
1624 aAlaLeuValAlaGluTyLeuAlaLeuLeuGlnLysAspGlnArgHisLeuP 1641  
120677 AGCACTGTTGCAAGATATCTCAACATGCTGAGAGATTTGTGCTACTTGC 120628  
||||| |||||

1641 rovalGlyCysValSer..... 1646  
|||||  
120627 CAATCGGTGCTGACCTTTAGCTGGGTGAGAACAGCTGACTTAAGT 120578  
1646 ..... 1646  
120577 ATGTGATGACGACAGTAGTGGGTGTATATAAATGCATCATACAT 120528  
1647 .... PheGlnAsnIleSerSerAsnValLeuGluGluSerAlaIleSerA 1662  
|||||  
120527 ACCGTTTCAACACATTTTCATCCAACGCTCGGAGGAGTCTGCTGTCTG 120478  
1662 spAspIleLeuSerProAspGluGluGlyPheCysSerGlyLysHisPhe 1678  
|||||  
120477 ATGACGCTGTCTCCGGAGGAGGAGGATCTGTGCTGGAAGTACTTC 120428  
1679 ThrGluLeuGlyLeuValGlyLeuLeuGluGlnAlaIleGlyTyrPheTh 1695  
|||||  
120427 AGCAGAGTGTGCTAGTGGGCTCTTGAGCAAGAGCTGCTCTCTTCAA 120378  
1695 rMet..... 1696  
|||||  
120377 CATGTACTGTGCTACATTTGTCTACACATGTGACAGAGTGGCTGTGTG 120328  
1696 ..... 1696  
120327 TCCCCATGGCTGTGGGGTTTATATGATTTGGCGTGTGTATAAACTGG 120278  
1697 ..... Gly.GlyLeuTyrGluAlaValAsnG 1705  
|||||  
120277 CCTACCAATGTGTGTTTACTGCGAGCGCTGCCATGTATGAAGCCATTAATG 120228  
1705 LuValTyrLysAsnLeuIleProIleLeuGluAlaHisArgAspTyrLys 1721  
|||||  
120227 AGGTTTACAAAGATTCGTCTGCCATCCATGACGCCAACAGACTTTAAG 120178  
1722 LysLeuAlaAlaValHisGlyLysLeuGluGlnAlaPheThrLysIleLeu 1738  
|||||  
120177 AAGTGGCTACAGTCATGGGAAGCTCAGAGAACTTTAAACAAAGCTTA 120128  
1738 His..... 1739  
|||||  
120127 CAACCAAGTAGTTTGACGGCTGATGACAGACAGATATCATTTTACTCT 120078  
1740 ..... GlnSerSer 1742  
|||||  
120077 GAAGTTCTCCATATGTTTGAATTTCTTTGTTTTCACCTTTCAGAGTTCA 120028  
1743 GlyTyr.Glu..... 1745  
|||||  
120027 GGATGGAGGATATGCTTATTTTAAAAAATGACTGATGAGCACTTACT 119978  
1745 ..... 1745  
119977 TTATTTATGTCACATTAACATTCCTGCTTCCGCTTATCTCCCCCT 119928  
1745 ..... 1745  
119927 TAATCCTTTTGAATGACAAATATATGGGTGATTTTAAACTATTTTA 119878  
1745 ..... 1745  
119877 TCCTAAAAATATGATTTTGTCTTAAAGCTGTGATTTAAATCATCAT 119828  
1746 ..... ArgValPheGlyThrTyrPheArg 1753  
|||||  
119827 CTTACCATCCCATCTTCAAGTGTAGAGAAATGTCGGGACCTATTTTCGG 119778  
1754 ValGlyPheTyrGlyAlaHisPheGlyAspLeuAspGluGlnGluPheVal 1770  
|||||  
119777 GTGGGTTTCTTGTGCTCTGTTTGGAGACCTGACAGACAGAGTTGT 119728

1770 ITyrLysGluProSerIlePheThrLysLeuAlaGluIleSerHisArgLeuG 1787  
|||||  
119727 CTACAGAGGACCATTCATCATAAATGCGACAGATCTCTCACAGACTTG 119678  
1787 Lu..... 1787  
119677 AGGCATGTTGACTCACCCAGCCTCAGAAAAAGAGTAATTAACGCTTCC 119628  
1788 ..... GluPheTyr 1790  
|||||  
119627 CAATTAATATAGCCTTTATATACGTGTGTTTGTGCTGCGACGAGTTTA 119578  
1790 rThrGluArgPheGlyAspAspValAlaGluIleIleLysAspSerTyrP 1807  
|||||  
119577 CTCAGAGCAGAGTTCCGAGATGACGTGTGTTAAATATATCAAGACATCCAAATC 119528  
1807 rovalAspLysSerLysLeuAspSerGlnLys..... 1817  
|||||  
119527 CAGTAGACAAAGAACTGATCCCAACAAAGTCCGTGATATTAGGC 119478  
1817 ..... 1817  
119477 CTCATGCCATAGGTGCTGCTCAGTTGTGTGTAATATTCTATAGAGTGT 119428  
1818 ..... AlaTyrIleGlnIleThrTyrValGluProTyrPheA 1830  
|||||  
119427 ATGTACTCTTTCAGGCTTACCTCCAGATCACCCTACCTGACCAACATTTCTCG 119378  
1830 sPThrTyrGluLeuLysAspArgValIthrTyrPheAspArgAsnTyrGly 1846  
|||||  
119377 ACACGATGAGCTTAAAGGAAAGAGTACCTACTTTGACAAAGACTTACAAAC 119328  
1847 LeuArgThrPheLeuPheCysThrProPheThrProAspGlyArgAlaNI 1863  
|||||  
119327 CTGCGTAACCTCATGTACTGTACTCTCCCTCAGCTCGAGCGGCGAGCACA 119278  
1863 sGlyGluLeuProGluGlnHisLysArgLysThrLeuLeuSerThrAsn 1880  
|||||  
119277 CCGCGAACTGAAAGAGCAGGTACAAACGTAACCACTCTCCAAAGCTGTCT 119228  
1880 fAlaPheProTyrIleLysThrArgIleArgValLysHisArgGluGlu 1896  
|||||  
119227 ACGCTTCCCTTACATCAAGACAGCATCAATGTATATCCACAGAGGAG 119178  
1897 Thr..... 1897  
|||||  
119177 GTGTGTTTGGAGAGCGGCAATCTCTAGGTGATGGATGCGTCCAGC 119128  
1898 ..... ValLeuThrProValG 1903  
|||||  
119127 AGGGGTATAAATTTGTTGCTGTTGTTCTCCAGATCAATTTCTGTTCCCATAG 119078  
1903 LuValAlaIleGluAspMetGlnLysLysThrArgGluLeuAlaPheAla 1919  
|||||  
119077 AGTGCGCATTTGAGGACATGCGAAGAGACACAGAGTGTGGCTTTGCC 119028  
1920 ThrGluGlnAspProProAspAlaLysMetLeuGlnMetValLeuGlnI 1936  
|||||  
119027 ACAAAACAGAGACCTGCGACAGACTCCAAGATGTTGACAGATGTGCTCCAGG 118978  
1936 ySerValGlyProThrValAsnGln..... 1944  
|||||  
118977 CTGCGTGGGACCACTGTCAACCAAGTGCTGTTTACACAGCTGTCAGA 118928  
1944 ..... 1944  
118927 TGTATATATCAATGAGTGTCAATATGCTATGCCCTTATTAATAGGCTT 118878  
1944 ..... 1944  
118877 TAAAGTCTGTCAGAGTGTATTATGTAGGCGCACATTCCTCGAGAA 118828  
1945 ..... Gly.ProLeuGluValAlaGlnVal 1952







1492 AsnPhcgluilegIyHisasnPhcAlaArgValIlysmetGlnValThrMe 1508  
|||||.....  
2234 AACTTGGATGGGAATTAACCTTGGCCAGGGTTAAATGCAGTAACAAT 2283  
1508 tSerLeuSerSerLeuValGlyThrThrGlnHisasnPhcSerGluGlnHisL 1525  
|||||.....  
2284 GTCATATCCCTCTTGGGGCACATCTCAGAAATTTTAATGAAGAATTCCT 2333  
1525 euArgArgSerLeuIlysthrIleLeuThrTyraIaGluGluAspMetGly 1541  
|||||.....  
2334 TAAGACGTCTCTAAAGACTATATTGACATATGCTGAAGAAGACTGTGAA 2383  
1542 LeuArgAspSerThrPheAlaGluGlnValGlnAspLeuMetPheAsnLe 1558  
|||||.....  
2384 TTGAGGGAAACACATTTCCGATCAGGTCCAGGATCTGGTTTTCATCT 2433  
1558 uHisMetIleLeuThrAspThrValIlysmetGlyHisGluGluAsp 1575  
|||||.....  
2434 CCATATGATTTCTTCTGATCTGAAATAAGAAACACAGAGAGATC 2483  
1575 roGluMetLeuIleAspLeuMetTyraIaArgIlyTyrglnGly 1591  
|||||.....  
2484 CTGAATGTTGATTCGATCTAATGTACAGAAATGCCAAGGGTTACCAACC 2533  
1592 SerProAspLeuAArgLeuThrThrTleuGlnHisMetIaGlyIlystHisAl 1608  
|||||.....  
2534 TCTCCAGATTCGCGATTGACCTGGTTCAGAAACATGGCAGCAGCAAGCTC 2583  
1608 aGluLeuGlyAsnHisAlaGluAlaIaGlnCysMetValHisIleAla 1625  
|||||.....  
2584 AGAAGCAAGCAATCATCTGTAAGCTGCACAGCTCTAGCTCACCACACAG 2633  
1625 IalLeuValAlaGlyIlyleuAlaLeuLeuGluAspGlnArgHisLeuPro 1641  
|||||.....  
2634 CACTTGTTCGGAATATTTGAGCATGCTGGAGGACCGGAATATCTTCTCT 2683  
1642 ValGlyCysValSerPheGlnAsnIleSerSerAsnValIleuGlnGly 1658  
|||||.....  
2684 GTGGGATGTGTAACATTCAGAAATATTCATCTAATGTTTGAAGAAATC 2733  
1658 IalAlIleSerAspAspIleLeuSerProAspGluGlnGlyPheCysSerG 1675  
|||||.....  
2734 TCGGCTCTCAGATGATGTGATCTCCAGATGAAGAAGATATCGCTCTG 2783  
1675 IyIstHisPheThrGluLeuGlyLeuValGlyLeuLeuGlnAlaIa 1691  
|||||.....  
2784 GAAATATCTTACTGAGTCAAGACTTGTGGGATTACTGAACAAGCAGCT 2833  
1692 GlyTyrrPheThrMetGlyGlyLeuTyrrGluAlaValAsnGlnValTyrrLy 1708  
::: |||||.....  
2834 GCTTCCTTCTCTAAGCTGGCATGTATGAAGCAAGTAAATGAAGTTTACAA 2883  
1708 sasnLeuIleProTleuGlnAlaHisArgSPtyrIlystHisLeuAla 1725  
|||||.....  
2884 AGTACTTATTCCTTATTCATGAAGCTAATCGGAATGCAAGAAGACTATCCA 2933  
1725 IalValHisGlyIlystLeuGlnGluAlaPheThrIlystIleMetHisGlnSer 1741  
|||||.....  
2934 CAATTCATGTGAACCTTCAAGAACATTCAGCAAAATTTGTCTCAAGAGT 2983  
1742 SerGlyTyrrGluArgValPheGlyThrTyrrPheArgValGlyPheTyrrG 1758  
::: |||||.....  
2984 ACTGGCTGGGAGCGATGTTGGACCTATTTTCGTGGTTTATATGG 3033  
1758 yAlaHisPheGlyAspLeuAspGluGlnGluPheValTyrrIlystHisPro 1775  
::: |||||.....  
3034 AACCAAGTTCGGGATTTGGATGAACAAGATTTGTTTACAAAGAGCGCTG 3083  
1775 erIleThrIlystLeuAlaGluIleSerHisArgLeuGlnGluPheTyrrThr 1791  
::: |||||.....  
3084 CAATTAACCAAACTTGCAGAGATATCTCAGAGATTGGAGGATTTTACGGA 3133

1792 GluArgPheGlyAspAspValAlaGluIleIleIlystAspSerTyrrProVa 1808  
|||||.....  
3134 GAAAGATTTGGAGAGAGATGTGGTTGAATTAATCAAGACTCTAATCCCTG 3183  
1808 IAspIlystSerIlystLeuAspSerGlnIlystAlaTyrrIleGlnIleThrTy 1825  
|||||.....  
3184 AGACAAGTGTAAATTAGATCTTAACAAGGCATATATTCGATTACCTTAG 3233  
1825 aIGluProTyrrPheAspThrTyrrGluLeuIlystAspArgValThrTyrrP 1841  
|||||.....  
3234 TGGAGGCATACCTTTGACACATATGAGATGAAGACAGAAATCACTATTC 3283  
1842 AspArgAsnTyrrGlyLeuArgThrPheLeuPheCysThrProPheThrPr 1858  
|||||.....  
3284 GACAAATAATTACATCTTCGTCGATCATGTACTGATACCTTTACTTT 3333  
1858 oAspGlyArgAlaHisGlyLeuLeuProGluGlnHisIlystArgIlystH 1875  
|||||.....  
3334 AGATGGCCCGCCATGGGGAACCTTCATGACAAATTCMAAAGGAAGCCA 3383  
1875 euLeuSerThrAspHisAlaPheProTyrrIleIlystThrArgIleArgVal 1891  
::: |||||.....  
3384 TTCTGACTACGTCTCATGCTTCTTCTTATATTAATAAAGGCTCAATGC 3433  
1892 CysHisArgGluGluThrValLeuThrProValGluValAlaIleGluAs 1908  
|||||.....  
3434 ACTCATAAAGAAAGATCATCTTAACACCAATTAAGTTGCTATTGAGGA 3483  
1908 pMetGlnIlystIlystThrArgGluLeuAlaPheAlaThrGlnGlnAspPro 1925  
|||||.....  
3484 CATGCAGAAAGAACACAGAGATGGCATTTGCCAACACATCCAGATCCCG 3533  
1925 roAspAlaIlysmetLeuGlnMetValIleuGlnGlySerValGlyProThr 1941  
|||||.....  
3534 CAGACCCCAAAATGCTTCAGATGTGCTCAGAGATCTGTAGGCACACCA 3583  
1942 ValAsnGlnGlyProLeuGlnValAlaGlnValPheLeuAlaGluIlePr 1958  
|||||.....  
3584 GTGAATCAGAGGGGCTTTGGAAAGTGGCCAGGTTTCTGTCTGAAATACC 3633  
1958 oGluAspProIlystLeuPheArgHisHisasnIlystLeuArgLeuCysPhe 1975  
::: |||||.....  
3634 TATGACCCAAAGCTCTTCAGACATCATATTAATGCGACCTCTCTTTA 3683  
1975 yAspPheCysIlystCysGluAspAlaLeuArgIlystAsnIlystAlaLeu 1991  
|||||.....  
3684 AAGATTTTACTTAAAGGTGTGAAGATGCCTTAAGAAAAAATTAAGAGCTTA 3733  
1992 IleGlyProAspGlnGluGluTyrrHisArgGluLeuGluArgAsnTyrrCy 2008  
|||||.....  
3734 ATTGGCCGGATCAAAAGAGATATCAAAAGGAAGTGGAGAGAACTATCA 3783  
2008 sArgLeuArgGluAlaLeuGlnProLeuLeuThrGlnArgLeuProGlnI 2025  
|||||.....  
3784 TCGCCTTAAGAGCGCCCTACAGCCACTGATCAACAAGAAAGCCCTCAGT 3833  
2025 euMetAlaProThrProProGlyLeu..... ArgAsnSerLeuAsnArg 2039  
|| ::: |||||.....  
3834 TATACAGGACGATATTGCTGTCACTGCGCACAGAGATCTTCAGTGA 3883  
2040 AlaSerPheArgIlystAlaAspLeu 2047  
|||||.....  
3884 ATGAGCCTTCGCAAAATGAGATCTC 3907

seq\_name: gb\_hlg:AC017375  
seq\_documentation\_block: 73515 bp DNA linear HTG 09-DEC-1999  
LOCUS AC017375  
DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered  
pieces.  
ACCESSION AC017375  
VERSION AC017375.1 GI:6553611  
KEYWORDS HTG; HTGS\_PHASE2.





```

39895 CATGCATTTCCACATCAGTTTAATGATGANTCATCCACCCTGGTGCACTT 39936
974 rValGlyLeuGluValIleThrArgValHisLysAspValGluLeuAlaG 991
||| ::::: :::::
39935 AGTTACAAACAAGTGGTGGCTACACACAGCAAGCAACCTAAATTACGCC 39886
991 LuHisLysAsnAlaSerLeuAlaPhePheLeuSerAspLeuLeuSerLeu 1007
:: ||||| :::::
39885 AGTCATTAATAAGCAAGTTAGCTTTTATATTGACATACATAAGCATC 39836
1008 ValAspArgGlyPheValPheSerLeuValArgAlaHisTyrLysGluVal 1024
::: ||||| :::::
39885 ATGCATGGGAGATTGTGTTGGCTCATCAAAACGTACCAAAAGTTCTT 39786
1024 LAlaThrArgLeuGlnSerSerProAsnProAlaIleLeuThrLeuAla 1041
: ::::: ||| :::::
39785 GATTTCCTAAATAAGCTTCAATACCGGAT.....CTAATGACATATA 39745
1041 rGMetGluPheThrArgIleLeuCysSerHisGluHisTyrValThrLeu 1057
::: ||||| :::::
39744 AAATAGACTTTTGAGAAATGTGTGTAGTCATGAGCATTTTGTAGCCTTA 39695
1058 AsnLeuPro.....CysCysProLeuSerProProAlaSerTr 1070
|||||
39694 AACTTCCCGTTTGGTACTTCATACACATGATGTAACAGCGCATGCAAGCC 39645
1070 oSerProSerValSerSerThrThrSerGlnSerSerThrPheSerSerg 1087
::: ||||| :::::
39644 CAGCCCAAGCAACAGTCGAGCAACAGCCAAACTTCTGGGTAAAGTAAA 39595
1087 InAla..... 1088
39594 AAACAATACACTTAATTTTACTATTGTGGCTCTTTTAAAGGGAT 39545
1089 .....ProAspProLysValThrSerMetPheGluLeuSergGlyProPh 1103
::: ||||| :::::
39544 CACTAGAGAGAGCCCTTCACGCC.....GACCTTAGTCAAGATTT 39505
1103 eArgGlnGlnHisPheLeuAlaGlyLeuLeuThrGluLeuAlaLeuAla 1120
||| ||||| :::::
39504 CCTTCACCAACATTTTCTAGTGGACTTGTCTGAGGACTTGGACGAC 39455
1120 LaLeuGlu...ProGluAlaGluGlyAlaPheLeuLeuHisLysLysAla 1135
::: ||||| :::::
39454 TAAAGGAGGAGCCAAACCCCAA.....TTACATGGAAGAAAGCA 39417
1136 lLeserAlaValHisSerLeuLeuCysGlyHisAspThrAspProArgTy 1152
||| :::::
39416 ATAAAGTGATTCGCAACSTGATGACATCTCACGATTTGGACCGCGGTA 39367
1152 rAlaGluAlaThrValLysAlaArgValAlaGluLeuTyrLeuProLeuL 1169
::: ||||| :::::
39366 TAGCGAAACGATGCACGTGCAAGGGTGGCTGCTTATATATATACCGCTAC 39317
1169 eUserIleAlaArgAspThrLeuProArgLeuHisAspPheAlaGluGly 1185
||||| :::::
39316 TTCTATATAGTAGATGATTCATTCCTCAGTTCGATCAGCACGGCCTTGAAG 39267
1186 ProGlyGlnArgSerArgLeuAlaSerMetLeuAspSerAspThrGluG 1202
::: :::::
39266 CAGGATGCTCTGCAGCAAAATCGGTCAACTGAGACTATCAAGGCGCAC 39217
1202 yGluGlyAspIleAlaGlyThrIleAsnProSerValAlaMetAlaIleAla 1219
::: ||||| :::::
39216 TCAAACTATTACAGCATCAACATTAATCCGGAAGTGCAATTTGGCATAT 39167
1219 laGlyGlyProLeuAlaProGlySerArgAlaSerIleSergGlnGlyPro 1235
::: ||||| :::::
39166 CGGGAAGTCTCCCTACTCTACTTGAATGACAACTTAAAAACAATTTG 39117
1236 ProThrAlaSerArgAlaGlyCysAlaLeuSerAlaGluSerArgTyTh 1252
||| ||||| :::::
39116 CCC.....CTCAGCTCGGAGAAATATACCGGCCA 39091

```

```

1252 rLeuLeuAlaCysValLeuTyrPValLeuLysAsnThrGluProAlaLeuL 1269
||||| :::::
39090 TTTATTGGTGTGTCTTCTGGGTGCTAAATAACTTGGAGCCCAACGTTT 39041
1269 eGluAlaArgTyrPAlaThrAspLeuThrLeuProGlnLeuGlyArgLeuLeu 1285
||| ||||| :::::
39040 TATATCGTGGCTTATAGATTTTAAGTCGCAATAGAGTGACCAAAATGCTT 38991
1286 AspLeuLeuTyrLeuCysLeuAlaAlaPheGluTyrLysGlyLysLys.. 1301
::: :::::
38990 CAAGGTGTTAAAGCTTGTGAAACCTTGCAGTACACAGGCAAAATAA 38941
1302 .....AlaPheGluArgIleAsnSerLeuThrPheLysLys.....S 1314
::: ||||| :::::
38940 CGTGGCCACCCTTAAAGCAACATATACAAACGTTTAGAGAAACTGGAT 38891
1314 eTLeuAspMetLysAlaArgLeuGluGluAlaIleLeuGlyThrIleGly 1330
||| ||||| :::::
38890 CTACGTGATGTTAAGGAAAGCTGGAAGAATGTATTAGAGGCCAAATTCA 38841
1331 AlaArgGlnGluMetValArgArgSerArgGluArgSerProPheGlyAs 1347
||||| :::::
38840 GCACGTTATGATTTATTAATGTCGAAAGATCGTAAT.....TC 38800
1347 nProGluAsnValArgTyrArgLys...SerValThrHisTyrLysGlnT 1363
: ||||| :::::
38799 AACAGAAATATCCGATGAGGAGAAAGACCAAGATGCCATATGCTCTCAT 38750
1363 hSerAspArgValAspLysThrLysAspGluMetGlnHisGluAlaLeu 1379
::: ||||| :::::
38749 ATGCGATGCTGTGGAAAAAGTGAGCATGAACCTGAATTAAGTCATTTG 38700
1380 ValGluGlyAsnLeuAlaThrGluAlaSerLeuValIleAspThrLe 1396
::: ||||| :::::
38699 ATCGAAGCTCTTTTGGCAACTGTAAGTTCGACTTTCGATTCGATACCT 38650
1396 uGluIleIleValGlnThrValMetLeuSergLysAlaArgLysValL 1413
||||| :::::
38649 TCAAATTAATTTGTCATGCTGCT.....GCCAACCTTATACAAATCTTC 38606
1413 eGluAlaValLeuLysValValLeuTyrSerLeuGlySerAlaGlnSer 1429
||||| :::::
38605 TTGGAAGTGTGTAAGAAGTCTTTCATCTCTTGTCCCGGAATCAAACTT 38556
1430 AlaLeuPheLeuGlnHisGlyLeuAlaThrGlnArgAlaLeuValSerLy 1446
::: ||||| :::::
38555 GTTTTGGCCCTGCAAAACTGTGTCCTCCCAAGCGCGTTAATCTTAA 38506
1446 sPheProGluLeuLeuPheGluGluAspThrGluLeuCysAlaAspLeu 1463
||||| :::::
38505 AATTCCAAACTTATATTATGATGACGAGACCGAATATATGTCGGATTTGT 38456
1463 yLeuAlaArgLeuLeuArgHisCysGlySerArgIleSerThrIleArgThr 1479
||||| :::::
38455 GCTTATTTTACTGAAACATTTGTGGGTCTCTATTGCGCAGGAATCGATCA 38406
1480 HisAlaSerAlaSerLeuTyrLeuLeuMetArgGlnAsnPheGluIleG 1496
::: ||||| :::::
38405 CAGGAGCTGCATCCTATATTACTATATGAGACAAATTTTGAATTTGG 38356
1496 yHis..... 1497
38355 AAATGTAAGTTTATTTTATTGTTAGTTTATTTAACTTAATTAATGCT 38306
1498 .....AsnPheAlaArgValLysMetGlnValThrMetSe 1509
||||| :::::
38305 CACTATTTCGATAGAACTTTCCCGAGTTTAAGTATGCAAGTGAAGATGTC 38256
1509 rLeuSerSerLeuValGlyThrThrGlnAsnPheSergLysGluHisLeuAla 1526
||||| :::::
38255 TTTAAGCTCCCTAGTTGGAACAAGTTCGCTTATTAGTGAACAATCTTAC 38206

```

1526 rgArgSerLeuYsrhrlleuThrTyrAlaGluGluAspMetGlyLeu 1542  
|||||.....  
38205 GCGGTCGCTTAACCGCTGCTTATGCTGATCCGATCCGACCTG 38156  
1543 ArgAspSerThrPheAlaGluGlnValGlnAspLeuMetPheAsnLeu 1559  
:|||||.....  
38155 CAGGATACTCTTTCTCGAACAACTACAGATTGCTTTTATCTGCA 38106  
1559 sMetIleLeuThrAspThrValLysMetLysGluHisGlnLysProG 1576  
|||||.....  
38105 TATGATCCCTGATGACTGTTAAATGAGAGATATCGAGAACCCGAG 38056  
1576 lMetLeuIleAspLeuMetTyrArgIleAlaArgGlyTyrGlnGlySer 1592  
|||||.....  
38055 AATGTGCTTACCTCATGATGCTATTCGCAAGGATACCAAAATAC 38006  
1593 ProAspLeuArgLeuThrTyrLeuGlnAsnMetAlaGlyLysHisAlaG 1609  
|||||.....  
38005 CCGATCTACGACTGACTGTGGAAAAATGCGTAAAAAAACCCGCA 37956  
1609 uLeuGlyAsnHisAlaGluAlaAlaGlnCysMetValHisAlaAlaL 1626  
| :|||||.....  
37955 GCGAGCAATTCACAGGAGACGCAATGTTATGATACATGCTGCTCT 37906  
1626 euValAlaGluTyrIleuAlaLeuLeuGluAspGlnArgHisLeuProVal 1642  
|||||.....  
37905 TAGTTCTCAATATCTTAGCATGTTGAGTCCACAAAACATTTGCTGTT 37856  
1643 GluCysValSerPheGlnAsnIleSerSerAsnValLeuGluGlnSerAl 1659  
|||||.....  
37855 GAGAGCTGAAGTTTTCACGAAATTTCTCCCAACACACTAATGAGATCGGC 37806  
1659 aIleSerAspAspIleLeuSerProAspGluGluGlyPheCysSerGlyL 1676  
| :|||||.....  
37805 CGATCTCGATGATGTCCTAAGTCCGCGCAAGATGCTATCTGCTAGGAA 37756  
1676 yHisPhehrgIleuGlyLeuValGlyLeuLeuGluGlnAlaAlaL 1692  
:|||||.....  
37755 ATCATTTCACTGAAACTGGGTTGAAGCCCTTCTGGAAGAACCTCCCAT 37706  
1693 TyrPheThrMetGlyGlyLeuTyrGluAlaValAsnGluValTyrLysAs 1709  
||| :|||||.....  
37705 TCCTTCAAGTTGCTGCAATGATGATGAGCAATGAGCAAGTCTACAAAT 37656  
1709 nLeuIleProIleLeuGluAlaHisArgAspTyrLysLysLeuAlaL 1726  
|||||.....  
37655 TCTAATACCATATGCGAGGCTAACAGAGATTTTCAAAAAGCTAAAGCAAG 37606  
1726 aHisGlyLysLeuGlnGluAlaPheThrLysIle..... 1737  
| :|||||.....  
37605 TTCATGCAAAATTGCGAGAGGCAATTAAATCCAACTACAGGTA 37556  
1738 .....MetHisGlnSerSerG 1743  
37555 ACAATATGTGTGAATTTTACCAACGGAATATATATATTATTAACA 37506  
1743 yLTPGluATcGvalPheGlyThrTyrPheArgValGlyPheTyrGlyAlaH 1760  
:|||||.....  
37505 GGGTAGAGAGGTTTGGAAACACTTCTGCTGCTTATGCGGAA 37456  
1760 lAspHeGlyAspLeuAspGluGlnGluPheValTyrLysGluProSerIle 1776  
|||||.....  
37455 AATTGGGGCTTGGATCAGCAGGAATTCATTTATAAGACCAACATG 37406  
1777 ThrLysLeuAlaGluIleSerHisArgLeuGlu..... 1787  
| :|||||.....  
37405 ACGAAGTTCGCCGAAATATTAGTGGCTTCAAGTATATATGCAAAATG 37356  
1788 .....Glu..PheTyr 1790  
37355 GAAAAATAGAACTATATCAATTTTGTTCACATACGTTAGAACCTTTAC 37306  
1791 ThrGluArgPheGlyAspAspValValGluIleIleLysAspSerTyrPr 1807

|||||.....  
37305 ACTGACAACTTCGACCGGACTCTGTGATATCATTAAGATTCCAAATAC 37256  
1807 oValAspLysSerLysLeuAspSerGlnLysAlaTyrIleGlnIleThrT 1824  
||||| :|||||.....  
37255 CGTTATATATATAGCTTGGATCCGATTAAGGCTTATCAATCAATTTACT 37206  
1824 yValGluProIlyrPheAspThrTyrGluLeuLysAspArgValThrTyr 1840  
|||||.....  
37205 ATGTTGAACCTTACTTGAACATATGAAATGCGTCATCGTAGACATAC 37156  
1841 pheAspArgAsnTyrGlyLeu..... 1847  
37155 TTTGAGCGGAATTCATATATAGATGATGATGAATTAACAGATAATTTA 37106  
| :|||||.....  
1848 .....ArgThrP 1850  
37105 AATCGAAATTTAAATTTATATTTTAAACATTTTGGTTTATGCAAGAGCT 37056  
1850 heLeuPheCys..ThrProPheThrProAspGlyArgAlaHisGlyLeu 1866  
| :|||||.....  
37055 TTTATATATGCGACGCTTTTACTTAAATAAGAAAGCAGCATGGCAACT 37006  
1866 uProGluGlnHisLysArgLysThrLeuLeuSerThrAspHisAlaPhe 1883  
| :|||||.....  
37005 AATGAGCAGGTAAACGGAACGATTTTGAACGCGGAGATACCTTC 36956  
1883 rGlyrIleLysThrArgIleArgValCysHisArgGluGluThrValLeu 1899  
||||| :|||||.....  
36955 CATAGCTAAAACTCGCATATATGTAATATAGCCGACACCAATATGCTCTG 36906  
1900 ThrProValGluValAlaIleGlu..... 1907  
| :|||||.....  
36905 GAACCGATTGAAGTGCAATTAAGTAAGTATGTTGTAATAATCCCAAA 36856  
1908 .....AspMetGlnLysLysThrArg 1915  
36855 GTTAATAAATTAACGTTATGCTAGATATTCAAAAAACATTTG 36806  
1915 luleuAlaPheAlaThrGluGlnAspProProAspAlaLysMetLeuGln 1931  
||||| :|||||.....  
36805 AATTGGCAGCTGCCAGATTCAGAGCAGCAGACCCCAAAATATGCA 36756  
1932 MetValLeuGlnGlySerValGlyProThrValAsnGlnGlyProLeuG 1948  
||||| :|||||.....  
36755 ATGCTACTCAGAGGATGATTTGGAACCACTGTTAACCAAGACCGATGA 36706  
1948 uValAlaGlnValPheLeuAlaGluIleProGluAspProLysLeu...P 1964  
| :|||||.....  
36705 AATGGCAGAGTGCTTCCAAATTTATCCGACGGAACAACCTGATACGA 36656  
1964 heArgHisHisAsnLysLeuArgLeuGlyPheLysAspPheCysLysLys 1980  
:|||||.....  
36655 CAAACACCAAAACAACTTCGTTGCTGCTTGGCAGTTTCAAAAGGCT 36606  
1981 CysGluAspAlaLeuArgLysAsnLysAlaLeuIleGlyProAspGlnL 1997  
||| :|||||.....  
36605 TGTGCGATGAGCTTTGAAAAGATCGCAATCTAATATCTTACAGATCAAA 36556  
1997 sGluTyrHisArgGluLeuGluGlnArgAsnTyrCysArgLeuArgGluAlaL 204  
| :|||||.....  
36555 AGATTACCAACGAGACTGGAACGTAAACACAGATGGTTCAATGAAAGAT 36506  
2014 euGlnProLeuLeuThr 2019  
| :|||||.....  
36505 TGACCTCCCTTATTTACT 36489  
seq\_name: gb\_in:AC008318  
seq\_documentation\_block: 166626 bp DNA linear INV 05-APR-2001  
LOCUS AC008318  
DEFINITION Drosophila melanogaster, chromosome 2L, region 21A-21A, BAC clone  
BACR05J20, complete sequence.



ACCESSION AC008318  
 VERSION AC008318.7 GI:13549312  
 KEYWORDS HTG.  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 166626)

Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H.,  
 Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,  
 Rogers, Y., An, H., Baldwin, D., Banazon, J., Beeson, K.Y., Busam, D.A.,  
 Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M.,  
 Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dreene, D., Farfan, D.,  
 Ferrera, S., Frise, E., Galle, R.F., Gary, N.S., George, R.A.,  
 Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,  
 Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,  
 McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,  
 Pacle, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,  
 Phanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,  
 Stapleton, M., Strong, R., Svitskas, R., Tector, C., Williams, S.M.,  
 Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Sequencing of Drosophila chromosome 2L, region 21A-21A

Unpublished

TITLE 2 (bases 1 to 166626)  
 JOURNAL Celniker, S.E., Agbayan, A., Arcaina, T.T., Baxter, E., Blazey, R.G.,  
 REFERENCE Butenoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
 AUTHORS Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
 Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,  
 Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, S.R., Mazda, P.,  
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacle, J.M., Park, S.,  
 Pfeiffer, B., Poon, L., Sequiera, A., Sethi, H., Shit, E.,  
 Svitskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
 Rubin, G.M.

Direct Submission

Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley  
 Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Apr 5, 2001 this sequence version replaced g1:6728945.

COMMENT Sequence submitted by:  
 Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory, MS 64-121  
 Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome  
 shotgun and from subclones of this BAC and its neighboring clones.  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our sequence  
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
 to [bdgpre@fruitfly.berkeley.edu](mailto:bdgpre@fruitfly.berkeley.edu).

FEATURES  
 source location/Qualifiers  
 1..166626  
 /organism="Drosophila melanogaster"  
 /strain="y: cn bw sp"  
 /db\_xref="taxon:7227"  
 /chromosome="2L"  
 /map="21A-21A"  
 /clone="BACR09J20 (DB41)"  
 /clone\_lib="RPCI-98 (Rosewell Park Cancer Institute  
 Drosophila melanogaster BAC library, partial EcoRI in  
 PBAC3.6)"

BASE COUNT 49737 a 32951 c 34017 g 49921 t  
 ORIGIN

alignment\_scores:  
 Quality: 4423.00 Length: 2260  
 Ratio: 2.881 Caps: 45  
 Percent Similarity: 67.920 Percent Identity: 43.761

alignment\_block:  
 us-09-736-968a-2 x AC008318 ..

Align seg 1/1 to: AC008318 from: 1 to: 166626

```

13  ILeasnaRgThrValAlaAlaGluValArgLySGluValSerArgGluAr 29
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
85623 ATCCCTAGCGACGTCGGCTGGCGATGTCGAAAAAATCTTCTGGTGC 85672
   29 gSerGlySerProHisSerSerArgArgCysSerSerLeuGlyValP 46
   : : : : : : : : : : : : : : : : : : : : : : : : : :
85673 TCTATCAAGGCAATGAGCACCACCATCATGTGTGTCTTCCGTAAAGC 85722
   46 roLeuThrgluValVal..... 51
   |||
85723 TATGTGTTAAATATACAAATTTTCTGTAGACATTCATTTTTCANA 85772
   52 .....GluProLeuAspPheGluAspValLeuLeuSerArgP 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
85773 GATCTCCCGACAGCCCTTGATTCGAAGAATTTCTATCCACATTA 85822
   64 roProAspAlaGluProGluProLeuArgAspLeuValGluPheProAla 80
   : : : : : : : : : : : : : : : : : : : : : : : : : :
85823 TGAACATCATCACCGGAGTCCCTGAACATATATTGGACTTCCGAA 85872
   81 AspAspLeuGluLeuLeuGluProArgGluCysArgThrgluP 97
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
85873 GGGGATGTAAGTCTTAAATCATTCGCCCAAAATCGAAGCTGTATCA 85922
   97 oGlyIleProLyAspGluLyLyLeuAspAla..... 107
   : : : : : : : : : : : : : : : : : : : : : : : : : :
85923 CGTGTCTCCCAATGAGATTTGTATGTACACCTAATCATACACAGC 85972
   108 .....GlnValArgAl 111
   |||
85973 TACATAGCTACAGTCTTTTAGTGTGATTTGCCCATCAGTCACAGA 86022
   111 AlaValGluMetTyrlleGluAspTrpValIleValHisArgArgTyrg 128
   : : : : : : : : : : : : : : : : : : : : : : : : : :
86023 ATGTGTTAACTGATAGCGCCCATGGAAGTGTGAATATGACACAGC 86072
   128 InTyrlLeuSerAlaAla..... 133
   : : : : : : : : : : : : : : : : : : : : : : : : : :
86073 GGCACCTTCCAGTCTGTGTATTCGCCGAGCGTATTCAGCCGGGAGAC 86122
   134 TyrSerProValThrThrAspThrGlnArgGluArgGlnTyrgLeuP 150
   ||| ||| : : : : : : : : : : : : : : : : : : : : : :
86123 ATAGTCTCTGGACATCCATTGCCAGTTGGCTTCAGTAAGCTTTGTACG 86170
   150 oArgGlnValPheGluGlnAspAlaSerGlyAspGluArgSerGlyProg 167
   : : : : : : : : : : : : : : : : : : : : : : : : : :
86171 ...TCGTCATTTGAGAGGCGTTCGCTATTAAGAGAGAGCTGCACCTC 86216
   167 LuAspSerAspAspSerArgArgGlySerGlySerProGluAsp..... 181
   : : : : : : : : : : : : : : : : : : : : : : : : : :
86217 CAAATTCGACATCCATTGCCAGTTGGCTTCAGTAAGCTTTGTACG 86266
   182 .....ThrProArgSerSerGlyValaSerSerIlePheAspLeuAr 195
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
86267 GACACTTTGACCCCGCGTGGTCTTGGCTAGC.....TTGATCTTCG 86310
   195 gAsnLeuAlaAlaAspSerLeuLeuProSerLeuLeuGlnArgAlaAlaP 212
   : : : : : : : : : : : : : : : : : : : : : : : : : :
86311 ACGTTCGGTATATGACCCCATATACCCCAATTTGCTTACAAATATACCTC 86360
   212 roGluAspValAspArgAsnGluThrLeuArgArgGlnHisArgPro 228
   ||| ||| : : : : : : : : : : : : : : : : : : : : : :
86361 CGGACATATATGATCAATCCATGACAGCAGCCGCGACCAAGATCGACAA 86410
   229 ProAlaLeuLeuThrLeuTyrlleProAlaProAspGluAspGluAlaValG 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
86411 ATGGCTCTTTTCTACGTATCCGGAATTCGAGAGGAGAAACATTTGA 86460
   245 uArgCysSerArgProGluProProArgGluHisPheGlyGlnArgTle 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
86461 ACGCCGCTGTCTCGCAGAGATTCAGTTGAACATATGGGTATCGCATTC 86510
   262 euValLyCysLeuSerLeuLyPheGluIleGluIleGluProIlePhe 278

```





```

90018 ATCGAAGGCTTTGGCACTGAGTTGCATTGCTTCGATACCT 90067
1396 ugiuileilevalaInthrValmetleusergluaIargIuserVal 1413
90068 TGAATAATATGTTGATCATGTGCT.....GCCAACCTTTATCAAAATCTTC 90111
1413 euGIyAlaValleuLysValValleuTySerleuGIySerAlaInser 1429
90112 TTGGAACTGTGCTAAAGGTGCTTCATTCCTTTGTCGCCGAATCATCT 90161
1430 AlaleuPheugInhisGlyleuAlaInthrGlnArgAlaValleuValSerly 1446
90162 GTTTTGCCCTGCAAAACTTTGTCCTCCCAACGCCGCTTAATCTTAA 90211
1446 spheProGluLeuPheGluGluAspThrGluLeuCyAlaAspLeu 1463
90212 ATTTCCAAACTTAATATTTGATGACGAGACCGATATATGCGGATTTGT 90261
1463 yslleuArgleuLeuArgHisCysGlySerArgIleSerThrIleArgThr 1479
90262 GTCTAATTTTACGAAACATTTGGGTCCCTAATTTGCCAGGAATGATCA 90311
1480 HisAlaSerAlaSerleuTyrlleuLeuMetArgGlnAsnPheGluIleGI 1496
90312 CAGGCACTGCATCACTATATTACTAATGACACAAATTTTGAATTTGG 90361
1496 y.His..... 1497
90362 AAATGTAAGTTTATTTATTTAGTTTATTTAACTAATTAATAGT 90411
1498 .....AsnPheAlaArgValLysMetGlnValThrMetSe 1509
90412 CACTTATTCGTATAGAACCTTGGCCGAGTTAAGATGCAAGTGACGATGTC 90461
1509 rleuSerSerleuValGIyThrThrGlnAsnPheSerGluGlnHisLeuA 1526
90462 TTTAAGTCCCTAGTTGGAACAAGTTGCGGTCTTACTAGACAAATCCTTAC 90511
1526 rArgSerleuLysThrIleleuThrTyrlaGIuGIuAspMetGIyLeu 1542
90512 GCCGTGGCTTAAACCGTCCCTGTATGCTGAATCCGACCTCGACCTG 90561
1543 ArgAspSerThrPheAlaGluGlnValGlnAspLeuMetPheAsnleuH 1559
90562 CAGGATATCTTTTCTCTGACACAGATACAGATTGCTTTTATATCTGCA 90611
1559 smetIleleuThrAspThrValLysMetLysGlnHisGlnGluAspProG 1576
90612 TATGATCTCTGACATACGTGTAATAAGTAAGATACAGGAAGACCCAG 90661
1576 lueMetleuIleAspLeuMetTyrlArgIleAlaArgGIyTyrlGlnGlySer 1592
90662 AAATGTGCTTGAACCTCATGAATGCTATTTGCCAAAGGATACCAAAATAAC 90711
1593 ProAspLeuArgleuThrTrPleuGlnAsnMetAlaGlyLysHisAlaGI 1609
90712 CCTGATCTAGACCTGACTTGCTGGTGAATAATGTGCTAAAAAACCGCGCA 90761
1609 uleuGIyAsnHisAlaGluAlaAlaGlnCysMetValHisAlaAlaAla 1626
90762 GCGAGCAAAATCACACGGAACACCACTGTGTATGTACATGCGCTTCCT 90811
1626 euValaGIuTyrlleuAlaLeuLeuGluAspGlnArgHisleuProVal 1642
90812 TAGTTCTGTAATATCTTACATGTTGAGTGCACAAAACATTTGCTGCTT 90861
1643 GLyCysValSerPheGlnAsnIleSerSerAsnValleuGluGluSerAl 1659
90862 GGAAGCTGTAAGTTTTCACGAATTTCTCCCAACACACTAATGAGTCGGC 90911
1659 aIleSerAspIleleuSerProAspGIuGluGlyPheCysSerGIyL 1676

```

```

90912 CGTATCGATGATGTGCTAAAGTCCCGGCAAGATGTAATCTGCTAGGA 90961
1676 yshIsPheThrGluLeuGlyleuValGlyleuLeuGluGlnAlaIaGIy 1692
90962 ATCATTTTCAAGAACTGGTGTGAAGGCTTGTGGAAGAGACCTCCAAAT 91011
1693 TyrPheThrMetGIyGlyleuTyrlGluAlaValAsnGluValTyrlLysAs 1709
91012 TCTTTCAAGTTGCTGGCATGTATGAAACAATGAAACGAAGTGTCAAAAT 91061
1709 nleuIleProIleleuGlnAlaHisArgAspTyrlLysLysleuAlaIaVal 1726
91062 TCTATATCCCAATATCCGAGGCTACACAGATTTTCAAAAGTACGAAG 91111
1726 aHisGIyLysleuGlnGluAlaPheThr..... 1735
91112 TTCAATGCGCAAAATTCGAGAGCATTTAATCGAATATCCCACTACAGTA 91161
1736 .....LysIle.MetHisGlnSerSerG 1743
91162 ACAATATTTGTAAATTTTACCACGCAAAATATATACATATTTATTAAC 91211
1743 lYrPrGIuArgValPheGIyThrTyrlPheArgValGIyPheTyrlAla 1759
91212 AGGTAAGAGATTTTGGAAACATACCTTTCGTGTGGCTTCTATGGCGGA 91261
1760 HisPheGIyAspLeuAspGIuGlnGluPheValTyrlLysGIuProSerI 1776
91262 AAATTTGGGACTTGTGATCAGCAGCAATTCATTTAATAAGCCACACAT 91311
1776 eThrLysleuAlaGluIleSerHisArgleuGlu..... 1787
91312 GACGAAGTTGCCGCAAAATATTAGTCGGCTTCAGGTATATTTGCCAATT 91361
1788 .....Glu.PheTy 1790
91362 GGAATAATATAGACTAATCAATTTGTTTCAACATACGTAGACTTTTA 91411
1790 rThrGIuArgPheGIyAspAspValValGIuIleLysAspSerTyrlP 1807
91412 CACTAAGCATTTGGACCGGACTGTGCAATATCATTAATAATTCACAATA 91461
1807 roValAspLysSerLysLysAspSerGlnLysAlaTyrlleGIuIleThr 1823
91462 CCGTTGATATTAATAGCTTGGATCCCGATTAAGGCTTACATTCAAATTA 91511
1824 TyrlValGIuProTyrlPheAspThrTyrlGluLeuLysAspArgValThrTy 1840
91512 TATGTTGAACCTACTTGTGAACAATATGAATGCGTCATCGTAGACATA 91561
1840 rPheAspArgAsnTyrlGlyLeu..... 1847
91562 CTTTAGCGGAATTTCAATATATAGATATGAATTAACAGATTAATT 91611
1848 .....ArgThr 1849
91612 AAATGCAATTTAAATATATATTTTAACATTTTGGTTATGCAAGAACG 91661
1850 PheleuPheCys.ThrProPheThrProAspGIyArgIleHisGIyLul 1866
91662 TTTTATATATGCGACCTTTTACTAAAAATGAAGAACACATGCGCGAAC 91711
1866 euPrGIuGlnHisLysArgGIyThrleuLeuSerThrAspHisAlaPhe 1882
91712 TAAATGACCAAGTGAAGCAAAAGATTTTGACACGCGCACTTCACTT 91761
1883 ProTyrlLysThrArgIleArgValCysHisArgGIuGluThrValle 1899
91762 CCATATGTAATAACTCGCATTAATGATTAAGCCAGACGAATAATGTCCT 91811
1899 uThrProValGIuValAlaIleGIu..... 1907
91812 GGAACCGAATTAAGTGAAGTAAGTAAGTAAGTATTAATATCCCAA 91861

```





```
90460 ....TCGCATTGTAAGAGGGTTCGCTTAATAAAGCGAGAGCTGCACCTC 90505
      :::: |||||::: :::: :::::||||
167 luaspSerAsnaspSerArgArgGlySerGlySerProGluasp..... 161
      :::::|||||::: :::::|||||
90506 CAAGTCTTCGCAATCGCAATGCAAGTTGGCTTCGTAACGCTCTGTACG 90555
      :::::|||||::: |||||
182 .....ThiProArgSerSerGlyAlaSerSerIlePheaspLeuAr 195
      |||||:::||||| |||||
90556 GACACTTTGACCCCGCGCTGGTCTTGGGCTAGC.....TTTGATCTTCG 90599
      :::::|||||::: |||||
195 gasnLeuAlaAlaaspSerLeuProSerLeuLeuGluArgAlaAla 212
      :::::|||||::: |||||
90600 ACGTTCGGTTAATGACCCACTAATACCAATTGCTTGACAAATGATACCTC 90649
      :::::|||||::: |||||
212 roGluaspValaspArgArgAsnGluThrLeuArgArgGlnHisArgPro 228
      |||||:::|||||::: |||||
90650 CGGAGCAATATTGCAATCCAAATGACACGCGCGCAAGATCGCAAA 90699
      :::::|||||::: |||||
229 ProAlaLeuLeuThrLeuTyProAlaProaspGluaspGluAlaValG 245
      |||||:::|||||::: |||||
90700 ATGGCTCTTTTCACCTGTATCCGATCTGAAACGCGAGAAACAATGTA 90749
      :::::|||||::: |||||
245 uArgCysSerArgProGluProProArgGluHisPheGlyGlnArgIle 262
      |||||:::|||||::: |||||
90750 ACGCGCGCTGCTCGAGAGATTCAGTTGAACAATATGGTCATCGCATTC 90799
      :::::|||||::: |||||
262 euValIysCysLeuSerLeuIysPheGluIleGluIleGluProIleP 278
      |||||:::|||||::: |||||
90800 AGGTTAACTGCTACAGCTTGCGCTGAGCGGAGAGTGGAGCCAAATTTT 90849
      :::::|||||::: |||||
279 GlyIleLeuAlaLeuTyArgPvalArgGluIlyLysLysIleSerGlu 295
      :::::|||||::: |||||
90850 GCATCGATGGCTATTATGATGTAAGAACGCGAAATAATTTCTGAAA 90899
      :::::|||||::: |||||
295 nPheTyPheaspLeuAsnSeraspSerMetIysGlyLeuLeuArgAla 312
      |||||:::|||||::: |||||
90900 TTTTCTATTTTACATGAAACGAGTACTTTAAACGCATGTTGTCAAGTC 90949
      :::::|||||::: |||||
312 IsGlyThrHisProAlaIleSerThrLeuAlaArgSerAlaIlePheSer 328
      |||||:::|||||::: |||||
90950 ATGTACAGTGGCGCGACATAAGTACCCAGAACCCCTACGCATATTCGAA 90999
      :::::|||||::: |||||
329 ValThrTyProSerProaspIlePheLeuValIleLysLeuGluLysVa 345
      :::::|||||::: |||||
91000 ATTTCATATTCGAGTAAAGATTTATTTCTGCTGATGCTGGAATAAAGT 91049
      :::::|||||::: |||||
345 IleuGlnGlnGlyaspIleSerGluCysGluProTyMetValLeuL 362
      |||||:::|||||::: |||||
91050 TTTTG...CAGGTCACATAAACAATTCCTGTCGACCCGAT.....CTCA 91090
      :::::|||||::: |||||
362 YsgIuValaspThrAlaLysAsnLysGluLysLeu...GluLysLeuArg 377
      |||||:::|||||::: |||||
91091 AAGAA.....GACAAAGATAATAATACCGAGAAAGTAAAA 91125
      :::::|||||::: |||||
378 LeuAlaIaGluGlnPheCysThArgLeuGlyArgTyArgMetProPh 394
      |||||:::|||||::: |||||
91126 TCCAAACCTGTGAGTACTGTGAACGACTAGCGCAATAATACCGAATGCCGT 91175
      :::::|||||::: |||||
394 eAlaThrThrAlaValHisLeuAlaAsnIleValSer..... 406
      |||||:::|||||::: |||||
91176 TGCATGACACAGAAATCTACCAACAATGTTTAAATGCGAATACTTTG 91225
      :::::|||||::: |||||
407 .....SerAlaGlyGlnLeuaspArgaspSer..... 415
      :::::|||||::: |||||
91226 AAAGCAAAAGAGCTGGCGCGCGCAACGGGATTCCTTTGGAATGCTTGT 91275
      :::::|||||::: |||||
416 .....AspSerGluGlyGluArgArgProAl 424
      :::::|||||::: |||||
91276 GGCAGACAGCTTGGCAGACGCGCAAGCTCAATAGCTTAGATGGAATAATC 91325
      :::::|||||::: |||||
424 aThrPthr.....AspArgArgArgGlyProGluasp..... 435
      :::::|||||::: |||||
91326 TTCCAAAAGTACTTTGATCAACTAAGCGAAAGCTAATGATATGACG 91375
      :::::|||||::: |||||
436 .....ArgAla 437
      :::::|||||::: |||||
91376 GAACACTTACACAGCGTGGTTCGTTGAACGAGAGAAAGCACCGCATTC 91425
      :::::|||||::: |||||
438 SerSerGlyAspAsp...AlaCysSerPheSerGlyPheArgProAlaTh 453
      |||||:::|||||::: |||||
91426 TGGTCCCGCAGATGATTTTGCAAACGTTGTAGAAAACCTTCGTCATATC 91475
      :::::|||||::: |||||
453 rLeuThrValThrAsnPhePheLysGlnGluAlaGluArgLeuSerAspG 470
      :::::|||||::: |||||
91476 TATTAACAGTACCAAGCTTCTTTAAACAAGAGCGACATTAATGAAGAACG 91525
      :::::|||||::: |||||
470 luAspLeuPheLysPheLeuAlaaspMetArgArgProSerSerLeuLeu 486
      |||||:::|||||::: |||||
91526 AGGATTTTGACAAAATTTTACAGAGATTAAACGACCGAGCTCACTATG 91575
      :::::|||||::: |||||
487 ArgArgLeuArgProValThrAlaGlnLeuLysIleaspIleSerProAl 503
      :::::|||||::: |||||
91576 AAGAGTACAAATGCATACCTGTTCTATTAATGAAATATCACCTTG 91625
      :::::|||||::: |||||
503 aProGluAsnProHisPheCysLeuSerProGluLeuLeuHisIleLysP 520
      :::::|||||::: |||||
91626 CGTCGAGGAAGCCAAAGAAATGCATTAACGCGACAGATTACCACTTAAC 91675
      :::::|||||::: |||||
520 roTyPProaspProArgGlyArgProThrLysGluIleLeuGluPhePro 536
      :::::|||||::: |||||
91676 CTCMAAGTCCACATATATGCGCCCACTTAAGGAATCTTAGAATTTGCG 91725
      :::::|||||::: |||||
537 AlaArgGluValTyArgAlaProHisThrSerTyArgAsnLeuLeuTyVa 553
      :::::|||||::: |||||
91726 CAATGGCGCATTTTACATCAATACATTAATAGCTATCCCAATCCCTATTGT 91775
      :::::|||||::: |||||
553 lTyPProHisSerLeuAsnPheSerSerArgGlnGlySerValArgAsnL 570
      |||||:::|||||::: |||||
91776 GTCACCAAGAGGCTAAATTTTCTTCGCGTACAGCTTCTGTGCAAAACA 91825
      :::::|||||::: |||||
570 euAlaValArgValGlnTyMetThrGlyGluaspProSerGlnAlaLeu 586
      :::::|||||::: |||||
91826 TTGCCGTTGCGCTCACTAATAGGCTGGAGAAACCGCAAAAGATCGGTC 91875
      :::::|||||::: |||||
587 ProValIlePheGlyLysSerSerCysSerGluPheThrArgGluAlaPh 603
      :::::|||||::: |||||
91876 AATGCCATCTACGGCAAGTCGCTTGTCCAAAATTTTCTACGAAGCAT 91925
      :::::|||||::: |||||
603 eThrProValValTyThrHisAsnLysSerProGluPheTyArgGluPheL 620
      |||||:::|||||::: |||||
91926 TACAGCTGTCATTTATCATACAAAATGCCCATCATCTTACGATGAATAA 91975
      :::::|||||::: |||||
620 ySLeuHisLeuProAlaCysValThrGluAsnHisLysLeuLeuPheThr 636
      |||||:::|||||::: |||||
91976 AATTCGCGTCCGCTGCATCAATAAAGCAGCATCACCATTTATTTGTTAC 92025
      :::::|||||::: |||||
637 PheTyThrHisValSerCysGlnProArgPro.....GlyThrAlaLe 650
      :::::|||||::: |||||
92026 ATTTCACAGTTTCATGTCAAAAGAACACACAGACTTCGACCGCTGCT 92075
      :::::|||||::: |||||
650 uGluThrProValGlyPheThrTyPheProLeuLeuGlnHisGlyArgL 667
      :::::|||||::: |||||
92076 AGAGACTCTTATCGGCTACACATGTTGCTGCTTACGAAAGTGAANAAC 92125
      :::::|||||::: |||||
667 euArgThrGlyProPheCysLeuProValSerValaspGlnProProPro 683
      :::::|||||::: |||||
92126 TTAAGTTTGGGAATTTATCTTCTGTATGATGAAATACACCGCGGAA 92175
      :::::|||||::: |||||
684 SerTyPThrValLeuThrProaspValAlaLeuProGlyMetArgTyArg 700
      :::::|||||::: |||||
92176 AATTCCTCATTTATACACCAAGATGTTCAACGCTGGAATTAATGAGCT 92225
      :::::|||||::: |||||
700 lAspGlyHisLysGlyValPheSerValGluLeuThrAlaValSerSerV 717
      :::::|||||::: |||||
92226 GGCATATCACAGAGCGGCTTTTCTATTATGTAAAGCAGTACCGCAA 92275
```







```
1824 TyrValGluProThrPheAspThrTyrGluLeuLysAspArgValThr 1840
|||||
95801 TAAAGTGAACCCCTTCTTGAACATATGAATGCGTCATCGTGAGACATA 95850
|||||
1840 rPheAspArgAsnTyrGlyLeu..... 1847
|||||
95851 CTTTGACGGCAATTTCAATATAGATATGATATGAATTAACAGATAATT 95900
|||||
1848 ..... 1849
|||||
95901 AAATGCAATTTTAAATATATTTAAATTTTACATTTTGGTTATGACAGACG 95950
|||||
1850 PheLeuPheCys..ThrProPheThrProAspGlyArgAlaHisGlyLuu 1866
|||||
95951 TTTTATATATGACCGCGCTTTACTTAAATGGAAGGCGCATGCGCAAC 96000
|||||
1866 uProGluGlnHisLysArgGlyThrLeuLeuSerThrAspHisAlaPhe 1882
|||||
96001 TAAATGACGCGTGTAAACGGAACGATTTTGACACCGCGCAATCCTTT 96050
|||||
1883 ProTyrIleLeuThrArgIleArgValCysHisArgGluGluThrValIle 1899
|||||
96051 CCATAGCTAAACCTCGCATTTATGTAATAGCCGACAGCAATAGTCTCT 96100
|||||
1899 uThrProValGluValAlaIleGlu..... 1907
|||||
96101 GGAACCGATGAGATGCGCAATGAAAGTAAGTATGTTGTAATATCCCAA 96150
|||||
1908 ..... 1914
|||||
96151 AGTTAATAATTAACCTTATTCCTAGATATATCAAAAACCAATTTG 96200
|||||
1915 GluLeuAlaPheAlaThrGluGlnAspProProAspAlaLysMetLeuG 1931
|||||
96201 GAATGCGACGTCGACGAAATCAAGAGCCAGAGCCCAAAATATTTGCA 96250
|||||
1931 nMeValLeuGlnGlySerValGlyProThrValAsnGlnGlyProLeuc 1948
|||||
96251 AATGAGTACTGAGGATGATGTGAAACCACTGTTAACCAAGACGATGG 96300
|||||
1948 lValAlaGlnValPheLeuAlaGluIleProGluAspProLysLeu... 1963
|||||
96301 AATAGGCGATGTCCTCTTCCATTTATCCGACGGAACAACACTGTACG 96350
|||||
1964 PheArgHisHisAsnLysLeuArgLeuCysPheLysAspPheCysLys 1980
|||||
96351 ACAAAACACCAAAACAACTTCGTTGCTCTTCGCGAGTTTCAAAAGC 96400
|||||
1980 sCysGluAspAlaLeuArgLysAsnLysAlaLeuIleGlyProAspGln 1997
|||||
96401 TTGTCCTGAGGCTTTGAAAAGATCGCAATCTTAATTCAGATCAAA 96450
|||||
1997 ysgLysTyrHisArgGluLeuGluArgAspTyrLysArgLeuGluAla 2013
|||||
96451 AAGATTACCAACGAACTGGAACGTACACAGATGCTTCATTGAACGA 96500
|||||
2014 LeuGlnProLeuLeuThr 2019
|||||
96501 TTGACCTCCCTTTATTAAT 96518
|||||
seq_name: gb_pr:BC008335
seq_documentation_block:
LOCUS BC008335 2825 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, clone IMAGE:3506202, mRNA, partial cds.
ACCESSION BC008335
VERSION BC008335.1 GI:14249910
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Plutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE
AUTHORS 1 (bases 1 to 2825)
          Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-2001) National Institutes of Health, Mammalian
          Gene Collection (MGC), Cancer Genomics Office, National Cancer
          Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
          USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
          Email: gcgaps-r@mail.nih.gov
          Tissue Procurement: DCTD/DTF
          cDNA Library Preparation: Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
          DNA Sequencing by: Genome Sequence Centre,
          BC Cancer Agency, Vancouver, BC, Canada
          Info@cgsc.bc.ca
          Steven Jones, Jennifer Asano, Ian Bosgel, Yaron Butterfield,
          Susanna Han, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
          Letlicia Hsiao, Martin Krzywnski, Reta Kutsche, Oliver Lee, Soo
          Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
          Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
          Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,
          Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
          George Yang, Scott Zuyderdun, Marco Marra.
FEATURES
source Clone distribution: MGC clone distribution information can be found
          through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
          Series: IRAL plate: 23 Row: d Column: 12.
          Location/Qualifiers
            1..2825
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:3506202"
              /tissue_type="kidney, renal cell adenocarcinoma"
              /clone_id="NIH_MGC_14"
              /lab_host="DH10B-R"
              /note="Vector: pOTB7"
            <1..2585
              /codon_start=3
              /product="unknown (protein for IMAGE:3506202)"
              /protein_id="AAH08335.1"
              /db_xref="GI:14249911"
              /translation="ORSRLASMDSPTEBEGDIAGINPVSAMATAGSPGASRASIT
              SCGPPTASRACGLSABSSRTLLACYLVTKTEPDLDRMAITDILPLDGLDLILY
              LCLAAFEYKGRKAFERINSLTEFKSLDMKARLEALIGTIGARQEVNRSRRSPGN
              PENVRRNKSTVTHKKQTSRDVKTQKDEMEHAEVGLAVDASLVVDLEIIVQVYM
              SEARESYLGAIVLVVLSGASQSAFLQHLGLTORALYSKPELLEFDETELCADLC
              RLRLRHGSRISITRTHASASLYLAKRONPEIGHNRAVKMVOYMSLSLVTGTONFS
              EEHLRSIKITLTYAFEDMGLRSTFAEOYDLMFIMHITIDTYMKRHOBDPMIL
              DIMYRLARGTGGSPDLRLTYLQNMACKNAHELGNHAFACOCVHAAALVAVETLALLEDH
              RHLPGCVSFQNTSSNVLESASISDILSPDEDEGFCGKHFTELGLVLELEQAGFT
              MGILYAVNVEVYNLIPILEAHNDYKRLAAVHGKLOEATPKIMHOSGWERVGYFR
              VGFYGAHFGDLEDEFEYKPSITKLAISHRLSELEETPFGDDVVEIKDNPYDKS
              KLSOKRAYIOITVEVPYFDYELKDDYEDRNRYGLRFTLFCPPFGDRAAGLEPDS
              HKRKTLSLTDHAPRYIKTRIRVCHRETYLTVLEVAIEEMOKKREPLAATGDDPDA
              KMLQNVYQSGVPTVNOGPLEVAOVFLAIPEDPKLFRHNHLRUCDFKCKCEDAL
              RKNKALIGDPQKYEARELERNYCRLEALQPLTORLPOLMAPTPPGLRNSLNRSFR
              KADL"
BASE COUNT 641 a 835 c 827 g 522 t
ORIGIN
alignment_scores:
  Quality: 4419.00 Length: 860
  Ratio: 5.144 Gaps: 0
  Percent Similarity: 99.884 Percent Identity: 99.767
alignment_block:
  US-09-736-968a-2 x BC008335 ..
Align seg 1/1 to: BC008335 from: 1 to: 2825
1188 GlnArgSerArgLeuAlaSerMetLeuAspSerAspThrGluGluGlu 1204
```

```

|||||
3 CAGCGGTCAAGACTGGCCCTCAATGCTTGACTGACACACAGAAAGCGAAGG 52
1204 YASPIIAIAGIYThrIleAsnProSerValAlaMetAlaIleAlaGlyC 1221
53 GGACATGTCGGGTHACATCAACCCCTCTGTGGCATGGCCATTGTGGTG 102
1221 LProLeuAlaProGlySerArgAlaSerIleSerGlnGlyProProThr 1237
103 GCCCCCTAGCCCGGTGGCTCCGGGGCAGCATCTCCAGGGGGCCACCAAG 152
1238 AlAserArgAlaGlyCysAlaLeuSerAlaGluSerSerArgThrLeu 1254
153 GCTTCTGCGCAGGCTGTGCCCTCTCTGTGAGTCAAGCCGAGCTTGCT 202
1254 uAlaCysValIleuThrValIleuYAsnThrGluProAlaLeuGln 1271
203 GCGCTGTGTCTGTGGGTCTGAAAACACCGAGCCGCGCTCTCCAGC 252
1271 rGTrAlaThrAspLeuThrLeuProGlnLeuGlyArgLeuAspLeu 1287
253 GCTGGCCACTGACTGACTGACCTCCCGAGCTGGAGCTGTGTGGACTTG 302
1288 LeuTyrlLeuYsIleuAlaIleAlaPheGluTyrlGlyGlyLysIleAlaPheGlu 1304
303 CTGTACTTGTGCTGTGGCTGTGAGTACAAAGGGAAAAAGCCTTGTA 352
1304 uArgIleAsnSerLeuThrPheLysSerLeuAspMetLysAlaArgL 1321
353 ACCGATCAACAGCCTCACAATTCAAAAAATCTGTGATATGAAAGCCCGGC 402
1321 euGlnGluAlaIleLeuGlyThrIleGlyAlaArgGlnGluMetValArg 1337
403 TAGAGGAAGCCATTCTGGTACCATCGAGCTGACAAAGAAATGTTCCGG 452
1338 ArgSerArgGluArgSerProPheGlyAsnProGluAsnValArgTrpArg 1354
453 CGAAGTGTAGAGAGAGCCCGTTTGGAAATCCAGAGAACGTGGCTGGGG 502
1354 glySerValThrHisTrpLysGlnThrSerAspArgValAspLysThrL 1371
503 GAAGAGGCTACACACTCGAAGCAAACTCAGACCCGCTGGACAACCA 552
1371 yAspGluMetGlnHisGluAlaLeuValGluGlyAsnLeuAlaThrGlu 1387
553 AGGATAAATGGAACACGAGGCTTGTGGAAAGGAACTGGCAACCGAG 602
1388 AlAserLeuValIleuAspThrLeuGlnIleIleValGlnThrValMet 1404
603 GCAAGCCTAGTGTCTGGACACACTGGAGATCATCTGCAGACGGTGAT 652
1404 tLeuSerGluAlaArgGluSerValIleuGlyAlaValIleuLysValValL 1421
653 GCTTTCAGAAAGCCCGGAGAGGCTCTTGGGGGAGTGCTGAAGGTTGTGC 702
1421 euTyrlSerLeuGlySerAlaGlnSerAlaLeuPheGlnHisGlyLeu 1437
703 TGTACAGCCTGGGCGAGTGGCCAAAGTGCCTTCTTGCAGCATGGCCCTG 752
1438 AlaThrGluArgAlaLeuValSerLysPheProGluLeuLeuPheGlnL 1454
753 GCCACCCAGAGGCGCTTGTGTCCAAGTTCGCCGAGCTGCTGTTCAGAGA 802
1454 uAspThrGluLeuCysAlaAspLeuCysLeuArgLeuLeuArgHisCysG 1471
803 GGACAGGAGCTGTGTGGACACTGTGCTGAGGCTCTACAGCACTGTGG 852
1471 LysArgGluSerThrIleArgThrHisAlaSerAlaSerLeuTyrlLeu 1487
853 GCAGCGCATCAAGCACCATCCGACACGACGCGCTCGCTGTAACCTG 902
1488 LeuMetArgGluAsnPheGluIleGlyHisAsnPheAlaArgValLysMet 1504
|||||

```

```

903 CTCATGCGACAAACTTCAGATCGGCCACAACTTTGCCCGTGAAGAT 952
1504 tGlnValThrMetSerLeuSerSerLeuValGlyThrThrGlnAsnPheS 1521
953 GCAGGTCAACATGTCTCTGTGCTCCCTGTGGGAGCAGCAGAACTTCA 1002
1521 euGlnGluHisLeuArgArgSerLeuLysThrIleLeuThrTyrlAlaGlu 1537
1003 GTGAAGACACACTGGCAGGTTCTACTCAAAACCATCTCACTATGTCTAG 1052
1538 GluAspMetGlyLeuArgAspSerThrPheAlaGluGlnValGlnAspLe 1554
1053 GAGGACATGGGGCTCGGGACAGCACCTTCCAGAGCAGGTCCAGGACT 1102
1554 uMetPheAsnLeuHisMetIleLeuThrAspThrValLysMetLysGluH 1571
1103 GATGTTCAACCTGCACATGATCTTGACGGACACGGTGTAAGTAGAAC 1152
1571 tSglnGluAspProGluMetLeuIleAspLeuMetTyrlArgIleAlaArg 1587
1153 ACCAGAGAGACCTGAGATGCTCATCGACCTCATGTACAGAAATTGCCCG 1202
1588 GlyTyrlGlnGlySerProAspLeuArgLeuThrTrpLeuGlnAsnMetAl 1604
1203 GGTACCAAGGGCTCACCGACCTTCGGCTGACTGTGTGCAAAACATGGC 1252
1604 aGlyLysHisAlaGluLeuGlyAsnHisAlaGluAlaIleAlaGlnCysMetV 1621
1253 CGGAAAGCACGGGAGCTGGGCCAACCCAGCGGCGCCCAAGTGCATGG 1302
1621 aHisAlaIleAlaLeuValAlaGluTyrlLeuAlaLeuLeuGluuAspGln 1637
1303 TGCACGCGCGCCCTCGTGGCTGTGAGTACTCGCCCTCTCGAGAGCACAC 1352
1638 ArgHisLeuProValGlyCysValSerPheGlnAsnIleSerSerArgAsn 1654
1353 GCGCACCTGGCCGTGGGCTGGCTTCTTCCAGAACATCTCATCAACGT 1402
1654 tLeuGlnGluSerAlaIleSerAspAspIleLeuSerProAspGlnGluG 1671
1403 GGTAAAGAGATCCGCATCTCCGAGACATCTCTGCGCCGACAGAGAGG 1452
1671 LysPheCysSerGlyLysHisPheThrGluLeuGlyLeuValGlyLeuLeu 1687
1453 GCTTCTGCTCCGGGAGACATTCACTGAGCTGGGGCTGGTGGGTCTCTG 1502
1688 GlnGlnAlaIleArgTyrlPheThrMetGlyGlyLeuTyrlGluAlaValAs 1704
1503 GAACAGGACGCGGCTACTTCAACATGGCGGGCTTACGAGGGGGTGA 1552
1704 nGluValTyrlLysAsnLeuIleProIleLeuGluAlaHisArgAspTyrl 1721
1553 TGAAGTCTACAAAGAACCTCATCCCATCTCTGGAAGCCACCGTACTACAA 1602
1721 yLysLeuAlaIleAlaValHisGlyLysLeuGlnGlnAlaPheThrLysIle 1737
1603 AGAAGCTGGCCGCGTGCACGCAAACTGCGAGGAGGCTTACACCAAAATC 1652
1738 MethIsglnSerSerGlyTrpGluArgValPheGlyThrTyrlPheArgVal 1754
1653 ATGCACCAAGATTCGGGCTGGAGCGGCTGTTCGGGACGGATTTTCCCGCT 1702
1754 tGlyPheTyrlGlyAlaHisPheGlyAspLeuAspGluGlnGluPheValT 1771
1703 GGGCTTACAGGCGCCCACTTCCGCTGACCTGAGACAGAGAGTTTGTGT 1752
1771 yTyrlGluProSerIleThrLysLeuAlaGluIleSerHisArgLeuGln 1787
1753 ACAAGAGACCATCATCAGACAGAACTGGCAGAGATCTCACACCGGCTGAG 1802
1788 GluPheTyrlThrGluArgPheGlyAspAspValValGluIleIleLysAs 1804
1803 GAGTTCTACACGAGAGATTGGCGGAGAGCTGCTTGAATATCAAGA 1852

```

1804 pSerTyrProValAlaAspLysSerLysLeuAspSerGlnLysAlaTyrIleG 1821  
1853 CTCAACCCCTGTGACAAAGTCCAAAGCTTGACTCACAAGGCTACATCC 1902  
1821 InIleThrTyrValGluProTyrPheAspThrTyrGluLeuLysAspArg 1837  
1903 AGATCAGGTATGTGAACCGTACTTGTATCATCAAGAGCTCAAGACCGG 1952  
1838 ValThrTyrPheAspArgAsnTyrGlyLeuArgThrPheLeuPheCysThr 1854  
1953 GTGACCTACTTGTACCCCAACTATGGGCTTCGCACATCTCTGTCTGCAC 2002  
1854 rProPheThrProAspGlyArgAlaHisGlyGluLeuProGluGlnHisL 1871  
2003 GCCGTTCACCCGCGATGGCGCGACACGGGAGAGCTCCGACCAACACA 2052  
1871 ysArgLysThrIleuSerThrAspHisAlaPheProTyrIleLysThr 1887  
2053 AGCTTAAGACGCTGTCTCAGACCGACACGCTTCCCTTACATCAGACT 2102  
1888 ArgIleArgValCysHisArgGluGluThrValLeuThrProValGluVa 1904  
2103 CGCATCCGCTGTGTGCCACCGGAGAGACGCTGCTACGCCAGTGGAGGT 2152  
1904 AlaIleGluAspMetGlnLysLysThrArgGluLeuAlaPheAlaThrG 1921  
2153 GGCCATGAGAGACATGCAAGAAAGACACGGGAGCTGGCTTGGCCACCG 2202  
1921 LuginAspProProAspAlaLysMetLeuGlnMetValLeuGlnGlySer 1937  
2203 AGCAGAGCCACCCAGATGCTAAGATGCTACAGATGCTTCAGGCTCT 2252  
1938 ValGlyProThrValAsnGlnGlyProLeuGluValAlaGlnValPheLe 1954  
2253 GTAGGGCCACCGCTGAACCAAGGCTCCCTGGAGGTGGCCAGAGTGT 2302  
1954 uAlaGluIleProGluAspProLysLeuPheArgHisAsnLysLeuA 1971  
2303 AGCAGAGATCCGGAAGACCCCAAGCTTTCGCGATCACAACAAATTGC 2352  
1971 rGluCysPheLysAspPheCysLysLysCysGluAspAlaLeuArgLys 1987  
2353 GGCTCTGCTTCAAGACTTCGCAAGAAATGTAGATGGCTCGGAGAA 2402  
1988 AsnLysAlaLeuIleGlyProAspGlnLysGluTyrHisArgGluLeuG 2004  
2403 AATTAAGGCCCTGATTGGCCGACCAAGAGATACACCGCTGAGTGA 2452  
2004 uArgAsnTyrCysArgLeuArgGluAlaLeuGlnProLeuLeuThrGlnA 2021  
2453 GCGCAACTACTGCAGCTGCGGAGGCTGTGACAGCCCTGCTTACCCAGC 2502  
2021 rGluProGlnLysMetAlaPheProThrProProGlyLeuArgAsnSerL 2037  
2503 GCGTGGCCCACTGATGGACCCACCCACCGGCTCAGAGAACTCTTGG 2552  
2038 AsnArgAlaSerPheArgLysAlaAspLeu 2047  
2553 AACAGAGCAAGTTTCGAAAGGACAGACCTC 2582  
seq\_name: gb\_pat:AX172980  
seq\_documentation\_block:  
LOCUS AX172980 4200 bp DNA linear PAT 03-JUL-2001  
DEFINITION Sequence 107 from Patent WO0142297.  
ACCESSION AX172980  
VERSION AX172980.1 GI:1459758  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 4200)

AUTHORS Lu, P., Garman, J. D. and Candia, A. F.  
TITLE Clasp-3 Transmembrane protein  
JOURNAL Patent: WO 0142297-A 107 14-JUN-2001;  
Arbor Vita Corporation (US)  
FEATURES  
source 1. 4200  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="polynucleotide fragment"  
BASE COUNT 1262 a 844 c 907 g 1187 t  
ORIGIN  
alignment\_scores:  
Quality: 4273.50 Length: 1270  
Ratio: 3.998 Gaps: 12  
Percent Similarity: 84.173 Percent Identity: 66.220  
alignment\_block:  
US-09-736-968a-2 x AX172980 ..  
Align seg 1/1 to: AX172980 from: 1 to: 4200  
854 PropProValThrValGlnAlaIleAlaThrLeuAlaArgGlySerGlyArgPr 870  
12 CCCGCTCGGTCCGACGGATATGATGATGATGATGATGATGATGATGATGATG 59  
870 calaSerLeuTyrLeuAlaArgSerLysSerLysSerLysSerLysSerLys 887  
60 .....GTAAATCGTCTCGAAGCCCTGATATATACCAATTCAG 96  
887 splLeuAlaValAlaProGlySerValAspAspGluValSerArgIleLeu 903  
11:.....:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| 905  
904 AlaSer..... 905  
147 GGGAGTAAGGGTTTAGATGCTCAATTCCTGGTTAACTGCTGTGTC 196  
905 ..... 905  
197 AAAAGTGGCCCATGGGGATCCACCCCAAGTCCAAAGTGCAGAAATCAACAC 246  
905 ..... 905  
247 AGGCTATGATGCAAGTGTATGCTATGCTTCGCACACAGACAGCTCA 296  
906 .....LysLeuLeuH 909  
297 AGTTTCTTCAAAACATTAAAGGAGCGCTTACCAACTAAAGCTTTTCA 346  
909 sglGluIleuAlaLeuGlnTTPValAlaSerSerAlaValAlaArgGluA 926  
347 CGAGGAGCTGGCTTTCAGTGGGTGTTTGGACGTGGACCGTTCGGGAT 396  
926 lalleLeuGlnHisAlaTrrPhePheGlnLeuMetValLysSerMet 942  
397 CAGCTTGCACACAGCGCTGCTTTTGAATATATGATAAGAGCAGC 446  
943 AlaLeuHisLeuLeuGlnArgLeuAspThrProArgLysLeuArg 959  
447 GTGACCACTTATATCTTAAATGATAACTTGAAGGCTCCAAAGGAAAAAGTGC 496  
959 gPrProGlyArgPheLeuAspArgLysAlaLeuValGlySerValG 976  
497 TTTTCCAGAACGTTTCATGATGATGATGATGATGATGATGATGATGATG 546  
976 lLeuGluValIleThrArgValHisLysAspValGluLeuAlaGlnHis 992  
547 CTAGTATATATGTTTACAGATTTTCAAGAGACACACAAATGTTGAGAGA 596  
993 LeuAsnAlaSerLeuAlaPhePheLeuSerAspLeuLeuSerLeuValAs 1009  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

597 CTCATACAGCCCTGCATCTCTTCATCATGATGTGTCGTGTTAAGCA 646  
1009 pArgGlyPheValPheSerLeuValArgAlaHisTyrLysGlnValAlaIar 1026  
647 CAGAGGATTTGTTTGTAGCCCTTATAAGTCTCTGCTATTAACAGAGTGTCT 696  
1026 hrrArgLeuGlnSerSerProAsnProAlaAlaLeuThrLeuArgMet 1042  
697 CAAGGCTTACTCATTAACCAATCCAGTGTCTGCTGCTTCAGGCTG 746  
1043 GluPheThrArgIleLeuCySerHisGlnHisTyrValThrLeuAsnLe 1059  
747 GATTTTTCAGCAATCATCTGCAGATCAGACACATATGTTACATTAAACTT 796  
1059 uProCySerProLeuSerProProAlaSerProSerProSerValSers 1076  
797 ACCCTCAGAGCTTACTTACTCCACCTGCATCTCCATCACCCTTGTTCTT 846  
1076 eDrThrThrSerGlnSerSerThrPheSerSerGlnAlaProAspProLys 1092  
847 CTGCACATCTCAGAGTCTGAGATTTTCTACGAATGTACAAAGTACCAAG 896  
1093 ValThrSerMetPheGluLeuSerGlyProPheArgGlnGlnHisPheLe 1109  
897 ATTCGAATATATGTTTGAATATTCGCTCTTCGCCACACAGCATTAATTT 946  
1109 uAlaGlyLeuLeuLeuThrGluLeuAlaLeuAlaLeuGluProGluAlaG 1126  
947 GGCAGGACTTGTGTTTACAGAGCTGGCTGTCTATTTAGACCCCTGATGCTG 996  
1126 LuGlyAlaPheLeuLeuHisLysLysAlaLeuSerAlaValHisSerLeu 1142  
997 AAGGACTGTTTGGATTCATTAAGAACTCATATGATGTACCAATTTTA 1046  
1143 LeuCySerGlnHisAspThrAspProArgTyrAlaGlnValThrValLysAl 1159  
1047 CTCGCAGTCAAGCTCAGACCCGCGGTACTGACCCCTCAGATTAAGGC 1096  
1159 dArgValAlaGluLeuTyrLeuProLeuLeuSerIleAlaArgAspThrL 1176  
1097 TCGAGTGGCCATGTTGTATCTACTCTGATTTGATATCATGATAAAGCTG 1146  
1176 euProArgLeuHisAspPheAlaGluGlyProGlyGlnArgSerArgLeu 1192  
1147 TACCTCAGCTGTATGATTTTACAGAACTCACAATCAACGAGGAAAGCCA 1196  
1193 AlAspSerMetLeuAspSerAspThrGluGlyGluAlaAspIleAlaGlyTh 1209  
1197 ATTTGTATACCACTGATGATTTATGAAGTGAAGCAGCAGCAGCAGCAGC 1240  
1209 rIleAsnProSerValAlaMetAlaIleAlaGlyLysProLeu..... 1223  
1241 GATTAAGCCAGACCGTGGCCATGCGCAATCGCAGACATCGGTCCTCAAC 1290  
1224 .....AlaProGlySerArgAlaSerIleSerGlnGlyProProThrAla 1238  
1291 TAAACAAGGCTGGCAGTTCCTCTCCTCAGCTCA.....ACGAGT 1328  
1239 SerArgAlaGlyCySerAlaLeuSerAlaGlnSerSerArgThrLeuLeuAl 1255  
1329 GGCAGGCAACACACTACCTTTTCAGCAGAAATCAAGTCAAGCCTTTGAT 1378  
1255 aCySerValLeuThrValLeuLysAsnThrGluProAlaLeuLeuGlnArgT 1272  
1379 CTGTCTACTTGGTTCCTCAAAAATGACAGATGAACAGTTCCTACACAAAT 1428  
1272 rPalThrAspLeuThrLeuProGlnLeuGlyArgLeuLeuAspLeu 1288  
1429 GGTTCACAGATCTCTCAGCTTTCAGACTTAACCGGCTATTAAGTCTGCTT 1478  
1289 TyrLeuCySerLeuAlaAlaPheGluTyrLysGlyLysLysAlaPheGluAr 1305  
1479 TATCTCTGTGTCTGTCTTGTGATTAAGGAAAGTGTGTTGAACG 1528  
1305 gIleAsnSerLeuThrPheLysLysSerLeuAspMetLysAlaArgLeu 1322  
1529 AATGATTAACCTTGACCTTTTAAGAAATCAAAACACATGAGACCAAGCTTG 1578  
1322 LuGluAlaIleLeuGlyThrIleGlyAlaArgGlnGluMetValArgArg 1338  
1579 AAGAACTATCTTGGAGCATAGGTGCCAGCAGAAATGTGATCGGCGA 1628  
1339 SerArg.....GluArgSerPro.....PheGlyAsnPr 1348  
1629 AGCCGAGGACACCTGCAGAGAAAGCCATCTGGAAGTCCCTTTGGAAGTCA 1678  
1348 oGluAsnValArgTyrArgLysSerValThrIstTrpLysGlnThrSerA 1365  
1679 AGAAATTTGAGAGTGGAGAGAAATATGACTCAGCTCAGGCTCAAAACACAG 1728  
1365 sPArgValAspLysThrLysAspGluMetGlnHisGlnAlaLeuValGlu 1381  
1729 AGAAGCTTGACAAATCAAGAGCAGAGATTGAACAGCAAGCACTGATGAT 1778  
1382 GlyAsnLeuAlaThrGluAlaSerLeuValAlaLeuAspThrLeuGluI 1398  
1779 GGAACCTGGCTTACAGAAACCAACTAATCATTTAGATTAACATTAGAGAT 1828  
1398 eIleValGlnThrValMetLeuSerGluAlaArgGluSerValLeuGlyA 1415  
1829 TGTGTGTGACACCTGTTTGTGTAACGAAATCCAAAGAGCATCTTGCTG 1878  
1415 lAlaValLeuLysValValLeuTyrSerLeuGlySerAlaGlnSerAlaLeu 1431  
1879 GAGTCTTAAAGTCTACTACACACATGAGCTGTAAACCAAGTGCAGT 1928  
1432 PheLeuGlnHisGlyLeuAlaThrGlnArgAlaLeuValSerLysPhePr 1448  
1929 TATCTACAAACACTGTTTGTGTAACAGAGAGCTTGTTCAAAAGTTTCC 1978  
1448 oGluLeuLeuPheGluGluAspThrGluLeuCySerAlaAspLeuCySerLeuA 1465  
1979 TCAACTCTTATTTGAAGAAGAGACAGACAGCTGCTGATTTATGCTCCA 2028  
1465 rGluLeuLeuArgHisCySerGlySerArgIleSerThrIleArgThrHisAla 1481  
2029 GGCCTTCTCGACACTGTAGCAGTACATCGGTACAAATCGGTCAACACCC 2078  
1482 SerAlaSerLeuTyrLeuLeuMetArgGlnAsnPheGluIleGlyHisAs 1498  
2079 AGTGCCTCCCTTACCTCTAATGAGCAGCAAACTTTGATGGGAATTA 2128  
1498 nPheAlaArgValLysMetGlnValThrMetSerLeuSerSerLeuValG 1515  
2129 CTTTCAGAGGTTAAATGACAGGTACCAATGTCATATCTCTGCTGGTGG 2178  
1515 lYThrThrGlnAsnPheSerGluGlnHisLeuAlaArgArgSerLeuLysThr 1531  
2179 GCACATCTCAGAAATTTAATGAGAATTTCTTAACAGCTTCTCTAAACACT 2228  
1532 lLeuLeuThrTyrAlaGluGluAspMetGlyLeuArgAspSerThrPheAl 1548  
2229 ATATGAGATATGCTGGAAGAGATCTGGAATTTAGAGGGAACAAATTTCC 2278  
1548 aGluGlnValAlaAspLeuMetPheAsnLeuHisMetIleLeuThrAspT 1565  
2279 TGATCAGGCTCCAGATCTGGTTTCAATCTCCATATGATTTCTTCTGATA 2328  
1565 hrrValLysMetLysGlnHisGlnLysProGluMetLeuIleAspLeu 1581  
2329 CTGTGAATAATGAGGAACACAGAGAGATCTCTGAAAGTGTGATTTGATGACTA 2378  
1582 MetTyrArgIleAlaArgGlyTyrGlnGlySerProAspLeuArgLeuThr 1598  
2379 ATGTACAGAAATTTGCCAAGGCTTACCAAGACCTCTCCAGATCTGCGATTGAC 2428







1443 TTGAGTATAAGCGAAAAAGTGTGACAGATGATGCTTGACCTTT 1492  
1312 LysLysSerLeuAspMetLysAlaArgLeuGluValAlaIleLeuGlyTh 1338  
1493 AAGAAATCAAAAGCATGAGAGCAAAAGCTTGAAAGCTATTCTTGGGAG 1542  
1328 rIleGlyAlaArgGlnGluMetValArgArgSerArg.....GluA 1342  
1543 CATAGGTCGACAGGCAGAAATGGTACGGGAGGACCGAGACAGCTCGAGA 1592  
1342 rGSerPro.....PheGlyAsnProGluAsnValArgTyrArg 1354  
1593 GAAGCCCATCTGGAGAGTGCCTTTGGAATCAAGAAATTTAGGTGGAGG 1642  
1355 LysSerValThrHisTyrPheGlnThrSerAspArgValAspLysThrLy 1371  
1643 AAAGTAAAGACTCACTGGCTCAAAACACAGAGAAGCTTGACAATCAAG 1692  
1371 sAspGluMetGluIleGluAlaLeuValGluGlyAsnLeuAlaThrGluA 1388  
1693 AGCAGAGATTGACACAGACGACTGATGATGAAACCTGGCTACAGAG 1742  
1388 LaSerLeuValValLeuAspThrLeuGluIleIleValGlnThrValMet 1404  
1743 CAACCTTAATCATTTAGATACATTAGAGATTGTGTTCCAGACCGTTTCT 1792  
1405 LeuSerGluAlaArgGluSerValLeuGlyValAlaLeuLysValAla 1421  
1793 GTAACGGAAATCCAAAGAGACGATCTTGTTGGAGTCTAAAGAGTCACT 1842  
1421 uTyrSerLeuGlySerAlaGlnSerAlaLeuPheLeuGlnHisGlyLeuA 1438  
1843 ACACAGCATGGCTGTACCCAAAGTGCAGTTTACTACAAACACGTTTGG 1892  
1438 LathrGlnArgAlaLeuValSerLysPheProGluLeuLeuPheGluGlu 1454  
1893 CTACACAGAGAGCCTTGTTCAAAAGTTTCTGTAACCTTATTGTAAGAA 1942  
1455 AsphrGluLeuCysAlaAspLeuCysLeuArgLeuLeuArgHisCysG1 1471  
1943 GAGCAGAGCGAGTGTGCTGATTTATGCTCAGGCTCTCCGACACTGTAG 1992  
1471 ySerArgIleSerThrIleArgThrHisAlaSerAlaSerLeuTyrLeuL 1488  
1993 CAGTAGCATGTGTACATACGCTACACCCACAGTCCCTCTTACCTAC 2042  
1488 euMetArgGlnAsnPheGluIleGlyHisAsnPheAlaArgValLysMet 1504  
2043 TAATGAGGCAAAACTTTGAGATTGGAAATACTTGGCAGGTTAAATG 2092  
1505 GlnValThrMetSerLeuSerSerLeuValGlyThrThrGlnAsnPheS 1521  
2093 CAGGTACCAATGTCCACTATCTCTCTGTTGGGCAATCTCAGAAATTTAA 2142  
1521 rGluGlnHisLeuArgArgSerLeuLysThrIleLeuThrTyrAlaGluG 1538  
2143 TGAAGAAATTTCTTAAGACGTTCTTAAGAGACTATATGACATATGCTGA 2192  
1538 LuAspMetGlyLeuArgAspSerThrPheAlaGluGlnValGlnAspLeu 1554  
2193 AAGATCTGGAATTGAGGAAACAACATTTCTCTGATCAGGTCCAGGATCTG 2242  
1555 MetPheAsnLeuHisMetIleLeuThrAspThrValLysMetLysGluH 1571  
2243 GTTTTCAATCTCATATGATTTCTTGTGATATCTGAAATGAAGAGACA 2292  
1571 sGlnGluAspProGluMetLeuIleAspLeuMetTyrArgIleAlaArgG 1588  
2293 CCAAGAGAGATCTGAATGTTGATGATCTTAATGTACAGAAATGCCAAG 2342  
1588 LysTyrGlnGlySerProAspLeuArgLeuThrTyrLeuGlnAsnMetAla 1604  
2343 GTTACCAAGACTCTCCAGATCTCGATTGACTGGTTGACAGAACATGGCA 2392

1605 GlyLysHisAlaGluLeuGluGlyAsnHisAlaGluAlaAlaGlnCysMetVa 1621  
2393 GGCAGAGACACACAGACAAACATCATGCTGAAGCTGCACAGTGTACT 2442  
1621 HisAlaAlaAlaLeuValAlaGluTyrLeuAlaLeuLeuGluAspGlnA 1638  
2443 CCACTCAGCAGCACTTGTCTGAATATTTGAGCATGTGAGAGCCGGA 2492  
1638 rHisLeuProValGlyCysValSerPheGlnAsnIleSerSerAsnVal 1654  
2493 AATATCTCTGTGGGATGTGTACATTTCAAGATATTTTCATCTAATGTT 2542  
1655 LeuGluGluSerAlaIleSerAspAspIleLeuSerProAspGluGlu 1671  
2543 TTGAGAAGAAATCTCGGCTCAGATGATGTGTCTCCAGATCAAGAAAG 2592  
1671 yPheCysSerGlyLysHisPheThrGluLeuGlyLeuValGlyLeuLeuG 1688  
2593 TATCTGCTGTGAAATACTTTACTGAGTCAGGACTTGTGGATTTACTGG 2642  
1688 LuGlnAlaAlaGlyTyrPheThrMetGlyGlyLeuTyrGlnAlaValAsn 1704  
2643 AACAGACAGCTGCTTCTCTCTATGGCTGGCATGTATGAACGAGTTAAT 2692  
1705 GluValTyrLysAsnLeuIleProIleLeuGluAlaHisArgAspTyrLy 1721  
2693 GAAGTTTACAAAGTACTTATTTCTATTCATGAAGCTAATCGGATGCANA 2742  
1721 sLysLeuAlaAlaValHisGlyLysLeuGlnGluAlaPheThrLysIleM 1738  
2743 GAACATATCCACAAATTCATGTAACTCAAGAGCATTCACCAAAATTG 2792  
1738 eHisGlnSerSerGlyTyrProGluArgValPheGlyThrTyrPheArgVal 1754  
2793 TTTATAGAGTACTGCTGCTGGAGCGGATGTTGGCACTATTGCTGTT 2842  
1755 GlyPheTyrGlyAlaHisPheGlyAspLeuAspGluGlnGluPheValTyr 1771  
2843 GGTTTTATGAAACCAAGTTCCGGGATTTGGATGAACAAAGATTTGTTA 2892  
1771 rLysGluProSerIleThrLysLeuAlaGluIleSerHisArgLeuGluG 1788  
2893 CAAGGAGCCTGCATAATCAACCAACTTCAGAGATATCTCACAGATTGGAG 2942  
1788 LuPheTyrThrGluArgPheGlyAspAspValValGluIleLysAsp 1804  
2943 GATTTTACGAGAAAGATTTGAGAGGATGTGTTGAAGTATCAAAAGAC 2992  
1805 SerTyrProValAspLysSerLysLeuAspSerGlnLysAlaTyrIleG1 1821  
2993 TCTAATCCGTGAGACAGTGAATTAATTAGTCTCAACAGGCTATATTTCA 3042  
1821 nIleThrTyrValGluProTyrPheAspThrTyrGluLeuLysAspArgV 1838  
3043 GATTAACCTATGTGAGCCCATCTTGGACACATATGAGATGAAGGACAGA 3092  
1838 aLThrTyrPheAspArgAsnTyrGlyLeuArgThrPheLeuPheCysThr 1854  
3093 TCACCTATTTCAGCAAAATTTACAATCTTGCGATTCATGACTGTACA 3142  
1855 ProPheThrProAspGlyArgAlaHisGlyGluLeuProGluGlnHisLy 1871  
3143 CCCTTACTTTGATGGCCGCCATGGGAGACTTCATGAACATTTCA 3192  
1871 sArgLysThrLeuLeuSerThrAspHisAlaPheProTyrIleLysThrA 1888  
3193 AAGGAAGACCAATTCGATACGTCATGCGCTTCTTATATTAATAACAA 3242  
1888 rGluArgValCysHisArgGluGluThrValLeuThrProValGluVal 1904  
3243 GGGTCAATGTACTCATTAAGAGAGATCATCTTAACCAATTTGAAGTT 3292

```
1905 AlatlleGIuaspmetGlnlyslYthrArgIuleuAlaIaphaIarhrgI 1921
|||||
3293 GCTATTGAGGACATGACAGAAAACACAGAGAGTTGGCATTTTGCACACACA 3342
|||||
1921 uGlnAspProProAspAlaIalysMetLeuGlnMetValleuGlnlySery 1938
|||||
3343 TCAGAGATCCGACAGACCCCAAAATGCTTACAGATGCTACTCCAGGATCTG 3392
|||||
1938 aIglYProthValaIasngInglyProLeuGlnIuAlaIaGlnValPheLeu 1954
|||||
3393 TAGGCACACAGATGATCAGAGGGGCTTTGGAGTGGCCGAGTTTCTG 3442
|||||
1955 AlaGluIleProGluAspProIysLeuPheArgHisAsnYlsLeuAr 1971
|||||
3443 TCTGAATTAAGTACTAGTACCCAAAGCTCTTCAGACATCATATAACCTGCG 3492
|||||
1971 gLeuCyshelYsAspPheCyshYsYsCygluAspAlaLeuArgIysa 1988
|||||
3493 ACTCTGCTTTAAAGATTTTACTTAAAGAGTGTGAAGATGCTTAAAGAAAA 3542
|||||
1988 snlysaIAlaIleuIleGlyProAspGlnlySGlnYrHisArgIuleuGln 2004
|||||
3543 ATTAAGAGCTTATGATGGCGGCTTCAAAAAGGATATCAAGAGGAATG... 3589
|||||
2005 Arg,AsnYrCyshArgLeuArgIuAlaIeugln,ProLeuLeuThrGln 2020
|||||
3590 GGGAACATGATCTGCGCTTAAAGAGCCCTACACACCT...AGATCAG 3636
|||||
2021 ArgLeuProGlnLeuMetAlaIProThr.....ProProGlyLeuArGAs 2035
|||||
3637 AAAGTCCCTGCTATCCAAAGCCAGTATGCTGTCCCTGCCACACAGAGA 3686
|||||
2035 nSerLeuAsnArgAlaSerPheArgIysAlaIeugln 2047
|||||
3687 TTCCTTACGTCGATGAGCTTTCGCAAAATGATGATCTC 3723
|||||
seq_name: gb_pat:AX172932
seq_documentation_block:
LOCUS AX172932 4144 bp DNA linear PAT 03-JUL-2001
DEFINITION Sequence 59 from Patent WO0142297.
ACCESSION AX172932
VERSION AX172932.1 GI:14597917
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Lu,P., Garman,J.D. and Candia,A.F.
TITLE Clasp-3 transmembrane protein
JOURNAL Patent: WO 0142297-A 59 14-JUN-2001;
Arbor Vita Corporation (US)
FEATURES
source
1..4144
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="human CLASP-3 cDNA"
197..3610
/note="human CLASP-3"
/codon_start=1
/protein_id="CAC43728.1"
/db_xref="GI:14597918"
/translation="MDRSCNRMSHTESSFLQTLTGLPTKLFHELEALQWVCSG
SVRESALQAMFEFFELMWVHLYFNDKLEAPKRSRPERFMDIALVSTIASDI
VSRFQDTEVERLNTSLAFPLNDLSVMDRGFVSLIKSCYKOVSKLSPNSVL
VSLRLDFRLTISHEHYVTLNPGSILPPASPSPSSATSSGSGFSFNODOKTAN
MELSYPRQHYLAGIVITELAVILDPBESLPELHKVIMVYNILSSHSDPRS
DPQIKARVAMLYPLIGITMETVPQIDFTEHNRGRFICADTDYSESGSMTSOT
VAMALAGTSVPQITRGSFLLTSGRQHTTTSASSSRSLICLLVMKMDETVLQK
WPTDLSVLQNLRLDLILYCVSGFEYKGRKYSERMNSLTFKSKDMRAKLEBALIGSI
GARQEWRSRGOLERSPGSAFGSOENLRMKDMTHMROMTERLDRKRAETEHALI
DGNLATEANLILDTLEIVYGVVSTESKESILGVLAVLHSMACNSAVYLVQHCPA
```

```
alignment_scores:
  Quality: 4273.00 Length: 1247
  Ratio: 4.039 Gaps: 11
  Percent Similarity: 84.844 Percent Identity: 67.121
alignment_block:
US-09-736-968a-2 x AX172932
Align seg 1/1 to: AX172932 from: 1 to: 4144
877 ArgSerLysSerIleSerSerSerAsnProAspIleuAlaIalProGln 893
|||||
11 CGTTCGAAAGCCTTAGTATAGCAATCCAGATATATCTGGGACATCCAC 60
893 YserValAspAspGluValSerArgIleLeuAlaSer..... 905
61 GTACACAGATGATGAGAGTTGCATCATATCGGAGTAAAGGTTTATGATC 110
905 ..... 905
111 GCTCAATTCCTGGGTTAACACTGGTGTCAAAAGCTGCCCATGGGAGA 160
905 ..... 905
161 TCCAACCCAGTCCAGATGACAGATTCACACAGGCTATGATCGAATTCG 210
905 ..... 905
211 TATCGTATGTCTTGGCAGACAGAGACGTCAAGTTCTTACAAACATTAA 260
906 ..... 906
261 CGGAGAGCTTACCACTAAAGAGCTTTTCACAGAGAGCTGGCTTTCGAG 310
916 TTPValValSerSerSerSerValValArgGluAlaIleLeuGlnHisAlaTr 932
|||||
311 TGGGTGTTTGGCAGTGGCAGCGTTCGGATGACGCTTTTCACACAGGCTG 360
932 pPhePheGlnLeuMetValLysSerMetAlaLeuHisLeuLeuG 949
|||||
361 GTCTTTTGAATTAATGATAAGAGCATGGTGCACCATTTATGACTTTA 410
949 YGlnArgLeuAspThrProArgGlyLeuArgPheProGlyArgPheLeu 965
|||||
411 ATGATAACTTGAGGCTCCAGAGAAAAGCGTTTTCACAGAGGTTTCATG 460
966 AspAspIleThrAlaLeuValGlySerValGlyLeuGluValIleThrAr 982
461 GATGACATTTGCACGCTTCTTCACAGACAGATGCTAGAGATATGTTTCACG 510
982 gValHisLysAspValGluLeuAlaGlnHisLeuAsnAlaSerLeuAlaP 999
|||||
511 ATTTCGAAGAGACACAAATGATGTTGAGAGCATATACAAACCTTGAT 560
999 hePheLeuSerAspLeuSerLeuValAspArgGlyPheValPheSer 1015
|||||
561 TCTTTCTCAATGATCTGTTGTTATGACAGAGGATTTGTTTTCAGC 610
1016 LeuValAlaGlnHisYrLysGlnValAlaIalThrArgLeuGlnSerSerPr 1032
```





```
/organism="Homo sapiens"
/ab_xref="taxon:9606"
/clone="MESAN1000101"
/cell_type="normal mesangial cells (NHMC56046-2)"
/clone_lib="MESANI"
/notes="cloning vector: PME18FL3-primary culture, normal.
mesangial cells"
54. .2279
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAB71042.1"
/db_xref="GI:16550750"
/translation="MEELSYPRROOHVLAIVLIDPAEGFLGKRVNMY
HILSHSDPRIDPDQIKARVAMLYPLGITIMETVPQIDLETINQGRICAT
DIYESSSMTISQVAAVIAIGTVSPQLRPSFLITSGRQHTTESAESSRLTICL
LVNKNADETVLQKMTDLVYLQNLRLDLVLCVSCFEYKGRKVEPNMSLPPKSK
DHRAKLEIAELIGISIGARQEMRSRGQLERSPSGASFOENLRMKDMHMQNTEK
LKSRAIEHEALIDGNLATVANLIIDTLEIVVQVSVTESKESIIGVLYKLHSM
ACNOAVYLQICFATROALVAFKPELLEETBECCALCLRLRHSSTIGTIRSHAS
ASVLYLRQNFQENNNFARVMQVMTSLSYGSONEENEPILRSKTLITVAEEDL
ELRTPPDQVODLYPNLHMTLSPTVYKMEQEDPEMLIDLMTRIAKGYOTSPDLRT
WIQNAGKHSRSHAEAOCLVSHALVAEYLSMLDRKTYLPGCVTFQNISSNVE
ESAVSDVAVSPDEKISGKFTESGLVGLLEQAAASFSAAGMEAVENEYKVLPIH
EANRDAKLSTHGKIQEAFSKIYHODGKRMFGYFVGVGKTFGDLDOEYVEEP
A1TXLAETSHRLEFGEFGEDEVVEYIKDSNPYDKCLDPNTHIFRLPMWSHTLTHM
R"
```

BASE COUNT 1046 a 640 c 711 g 930 t  
ORIGIN

alignment\_scores:  
Quality: 3626.50 Length: 982  
Ratio: 4.130 Gaps: 7  
Percent Similarity: 89.409 Percent Identity: 71.996

alignment\_block:  
US-09-736-968a-2 x AK055905 ..

Align seg 1/1 to: AK055905 from: 1 to: 3327

1079 SerGlnSerThrPheSerSerGlnAlaProAspProLysValThrSe 1095  
|||||  
3 TCTCAGAGTCTGGATTTCTACGAATGTACAGACCAAAAGATTGCGAAA 52  
1095 rMetPheGluLeuSerGlyProPheArgGlnGlnHisPheLeuAlaGlyL 1112  
|||||  
53 TATGTTGAATATCCGTGCTTTCCGCCACAGCAATTTTGGCGAGC 102  
1112 euLeuLeuThrGluLeuAlaLeuAlaLeuGluProGluAlaGluGlyAla 1128  
|||||  
103 TTGTGTTAAGAGAGGTGCTGTCATTTTAGACCTCATGCTGAGAGACG 152  
1129 PheLeuLeuHisLysLysAlaHisSerAlaValHisSerLeuLeuGlyScl 1145  
|||||  
153 TTTGGATTGCAATAGAAAGTCATCAATATGATACCAATTTACTCTCCG 202  
1145 yHisAspThrAspProArgTyrAlaGluAlaThrValLysAlaArgVala 1162  
|||||  
203 TCACGACTCAGACCCGCGTACTCTGACCTCAGATAAAGCGTCGAGTGG 252  
1162 LagLLeuTyrLeuProLeuLeuSerLLeaLArgAspThrLeuProArg 1178  
|||||  
253 CCATGTTGTATCTACCTCTGATTGATTAATCATGGAACCTGACTCTCAG 302  
1179 LeuHisAspPheAlaGluGlyProGlyGlnArgSerArgLeuAlaSerSe 1195  
|||||  
303 CTGTATGATTTTACGAAACTCAACAAATCAACGAGAAACCAATTGTAT 352  
1195 tLeuAspSerAspThrGlnGlyGluGlyAspLLeaLArgLysThrLeuAsp 1212  
|||||  
353 AGCCACTGATGATTTATGAAAGTAGAGC.....GGAAGATCATCAAGC 396  
1212 roSerValAlaMetValAlaLagLysProLeu.....Ala 1224

```
.....  
397 AGACCGTTGCCATGCATCGCAGGACATCGCTCCCTCACTAACAAG 446  
1225 ProGlnSerArgAlaSerLLeuSerGlnGlyProProThrAlaSerArgAl 1241  
|||||  
447 CTTGGCAGTTTCTCTCTCACGCTA.....ACAGAGGGAGGCA 484  
1241 agLysAlaLeuSerAlaGluSerSerArgThrLeuLeuAlaCysValL 1258  
|||||  
485 ACAACACTACCTTTTCAGACGAATCAAGTCGAAAGCCTTTGATCTGTAC 534  
1258 euThrValLLeuLysAsnThrGluProAlaLeuLeuGlnArgThrPlalThr 1274  
|||||  
535 TTTGGGTTCAAAAATGCGATGAAACAGTTCTTCAGAAAGGTTTACA 584  
1275 AspLeuThrLeuProGlnLeuGlyArgLeuLeuAspLeuLeuTyrLeuCy 1291  
|||||  
585 GATCTCACTCTTGCAGCGTAACCGCTATTAGATCTGCTTATCTCTCG 634  
1291 sLeuAlaAlaPheGluTyrLysGlyLysAlaPheGluArgLLeaAsn 1308  
|||||  
635 TGTCTCTGCTTGAATTAAGGAAAGAGTTTGACCAAGATGAATA 684  
1308 euLeuThrPheLysLysSerLeuAspMetLysAlaArgLeuGluAla 1324  
|||||  
685 GCTTGACCTTTAAGAAATCAAAAGACATGAGCAAAAGCTTGAAAGAGCT 734  
1325 tLeuGlnThrLLeuGlyAlaArgGlnGluMetValArgArgSerArg 1340  
|||||  
735 ATCTCTGGAGCATAGGTGCGACGCAAGAAATGTTAGGCGGCAAG 784  
1341 .....GluArgSerPro.....PheGlyAsnProGluAsnV 1351  
|||||  
785 ACAGCTCGAGAGAAAGCCCATCTGGAAGTCCCTTGGAAATGCAAGAAAT 834  
1351 alArgThrPArgLysSerValThrHisThrLysGlnThrSerSerArgVal 1367  
|||||  
835 TGAAGTGAAGGAAAGATATGATCACTGCGCTCAAAACCAAGAAAGCTT 884  
1368 AspLysThrLysAspGluMetGlnHisGluAlaLeuValGluGlyAsnLe 1384  
|||||  
885 GACAAATCAAGACAGCAGATTTGAACACGAAGCACTGATGATGAGAAACT 934  
1384 uAlaThrGluAlaSerLeuValAlaLeuAspThrLeuGlnLLeuValG 1401  
|||||  
935 GCTTACAGAAACCAACCTAATCATTTAGATACATTAGCATTTGTTTC 984  
1401 tThrValMetLeuSerGluAlaArgLLeuSerValLeuGlyAlaValLeu 1417  
|||||  
985 AGACCGTTCTGTACGGAATCCAAAGAGACATCTCTGTGGAGTGCCTA 1034  
1418 LysValValLeuTyrSerLeuGlySerAlaGlnSerAlaLeuPheLeuG 1434  
|||||  
1035 AAAGTCTACTACACAGCATGCGCTTAACCAAGAGCGATTATCTATA 1084  
1434 nHisGlyLeuAlaThrGlnArgAlaLeuValSerLysPheProGluLeuL 1451  
|||||  
1085 ACACTGTTTGTACACAGAGACCTTGGTTCAAGTTTCTCGAAGCTCT 1134  
1451 euPheGluGluAspThrGlnLeuLeuCysAlaAspLeuCysLeuArgLeu 1467  
|||||  
1135 TATTTGAAGAGAGACAGACAGCACTGTGCTGATTATGACCTCAGGCTTCTC 1184  
1468 ArgHisCysGlySerArgLLeuSerThrLLeuArgThrHisAlaSerAlaSe 1484  
|||||  
1185 CGACAGCTGACGAGTACGCTGATACATACGCTACACAGCCAGGCGCTC 1234  
1484 rLeuTyrLeuLeuMetArgGlnAsnPheGluLLeuGlyHisAsnPheAla 1501  
|||||  
1235 CCTTACTACTATATGAGCAAAACCTTGAGATGGGAATTAACCTTGCCA 1284  
1501 rGValLysMetGlnValThrMetSerLeuSerSerLeuValLysThrThr 1517  
|||||
```

1285 GGGTTAAATGCAAGTAAACATGTCATCTCTTGGTGGCACATCT 1334  
1518 GlnAsnPhseSerGluGlnHisLeuArgArgSerLeuLysThrIleLeuThr 1534  
1335 CAGAAATTTAAAGAAATCTCTTAAGACGTTCTCTAAAGACTATATGAC 1384  
1534 rTyraIaGluGluAspMetGlyLeuArgAspSerThrPheIaGluGln 1551  
1385 ATATGCTGAAGAAGATCTGGAAATGAGGAAACAACTTCCCTGATCAGG 1434  
1551 aIaGlnAspMetPheAsnLeuHisMetIleLeuThrAspThrValLys 1567  
1435 TCCAGATCTGGTTTCAATCTCCATATGATCTTCTGATACGTGAAA 1484  
1568 MetLysGlnHisGlnGluAspProGluMetLeuIleAspLeuMetThr 1584  
1485 ATGAAGAACACACAGAGATCTCGAAATGTTGATGATCTAATGTACAG 1534  
1584 gTleAlaArgGlyThrGlnGlySerProAspLeuArgLeuThrTrpLeu 1601  
1535 AATTTGCCAAGGTTTACCAGACCTCTCCAGATCTCGACTGACCTGGTTC 1584  
1601 InAsnMetIaGlyLysHisIaGluLeuGlyAsnHisIaGluAla 1617  
1585 AGAACATGGCAGGCAACACTCAGAACGAAGCATCATCTGAAAGCTGCA 1634  
1618 GlnCysMetValHisAlaAlaLeuValAlaGluThrLeuAlaLeu 1634  
1635 CAGTGTCTAGTCCACTCAGACGACTTGTGCAATATTTAGCGATGCT 1684  
1634 uGluAspGlnArgHisLeuProValGlyCysValSerPheGlnAsnHis 1651  
1685 GAGAGACCGGAAATATCTCTCTGGGATCTGTAACTTCAGAAATATT 1734  
1651 eSerAsnValLeuGluGluSerAlaIleSerAspSpilleuSerPro 1667  
1735 CATCTATGTTTGAAGAATCTGCGGTCTCAGATGATGGTATATCTCA 1784  
1668 AspGlnGluGlyPheCysSerGlyLysHisPheThrGluLeuGlyLeu 1684  
1785 GATGAAGAGGATCTGCTGTGAAATACTTTACTGACGACGACTGTG 1834  
1684 IglLeuLeuGluGlnAlaIaGlyThrPheThrMetGlyGlyLeuThr 1701  
1835 GGGATTCTCGAACACACGCTGCTCTCTATGCGTGCATGTATG 1884  
1701 IuaIaValAsnGluValThrLysAsnLeuIleProIleLeuGluAlaHis 1717  
1885 AAGCAGTTAATGAAGTTTACAAAGTACTTATCTTCAATGAAAGCTAAT 1934  
1718 ArgAspTrpLysLysLeuAlaIaValHisGlyLysLeuGlnGluAla 1734  
1935 CGGGATGCAAAAGAACTATCCACAAATTCATGTAACCTCAAGAGCAT 1984  
1734 eThrLysIleMetHisGlnSerSerGlyTrpGluArgValPheGlyThr 1751  
1985 CAGCAAAATGCTCATCAGGATGCT....AAGCGATCTTTGGCACCT 2028  
1751 yPheArgValAlaGlyPheThrGlyValHisIlePheGlyAspLeuAspGlu 1767  
2029 ATTTTCGCTGTTTATGAGACCAAGTTCCGGGATTTGAGAGAACAA 2078  
1768 GluPheValThrLysGluProSerIleThrLysLeuAlaGluIleSer 1784  
2079 GAATTTGTTTACAAGAGCTCGATAACCAACTGACAGATATCTCA 2128  
1784 sArgLeuGluGluPheThrGluArgPheGlyAspAspValAlaGlu 1801  
2129 CAGATTGAGGATTTTACGAGAAAGATTTGAGAGAGTGGTTGAG 2178  
1801 IleIleLysAspSerThrProValAspLysSerLysLeuAspSerGlnLys 1817  
2179 TAATCAAAAGACTTAATCTGTAGACAAAGTGAATTAATTAATGATCAAC 2227

1818 AlaThrIleGlnIleThrThrValGluProThrPheAspThrTyroIle 1834  
2228 GCATATATTCAGATTACTATGTGGACCATCTTTGACACATATAGAT 2277  
1834 uLysAspArgValThrTyPheAspArgAsnTrpGlyLeuArgThrPhe 1851  
2278 GAAGACAGAGATCCCTATTTCCGCAAAATTTACAACTTCTGTCGATTC 2327  
1851 euphneCysThrProPheThrProAspGlyArgAlaHisGlyGluLeuPro 1867  
2328 TGTACTGTACACCTTACTTACATGTAGATGCCGTGCCATGGGGAATTCAT 2377  
1868 GluGlnHisLysArgLysThrLeuLeuSerThrAspHisIaPheProTy 1884  
2378 GAACAAATTCAAAGAGACCATCTGTACATACGTCATCTCTCTCTTA 2427  
1884 rIleLysThrArgIleArgValCysHisArgGluGluThrValLeuThr 1901  
2428 TATTAAACAAGGCTCAATGTCACTCACTAAAGAAAGAGATCATCTTAAC 2477  
1901 roValGluValAlaIleGluAspMetGlnLysLysThrArgGluLeuAla 1917  
2478 CAATTGAAGTGTCTATTGAGGACATCGAGAAAAGACACAGAGTTGCA 2527  
1918 PheAlaThrGluGlnAspProProAspAlaLysMetLeuGlnMetVal 1934  
2528 TTTGCAACATCAGAGATCCCGCAGACCCCAAAATGCTTCAGATGTACT 2577  
1934 uGlnGlySerValGlyProThrValAsnGlnGlyProLeuGluValAla 1951  
2578 CCAGGATCTGTAGGACACACAGTGAATCAGGGCCCTTTGGAAGTTGCC 2627  
1951 InValPheLeuAlaGluIleProGluAspProLysLeuPheArgHisHis 1967  
2628 AGGTTTCTGTCTGAATTAACCTAGTAGACCAAGCTCTTCAGACATAT 2677  
1968 AsnLysLeuArgLeuCysPheLysAspPheCysLysLysCysGluAspAl 1984  
2678 AATAAAGTCGACCTGCTTTAAAGATTTTACTTAAAGGTGTGAAGATGC 2727  
1984 AleuArgLysAsnLysAlaLeuIleGlyProAspGlnLysGluThrHis 2001  
2728 CTTAAGAAAATAAGACTTAATTTGGCCGAGTCAAAAGAGTATCAAA 2777  
2001 rGluLeuGluArgAsnTrpCysArgLeuArgGluAlaLeuGlnProLeu 2017  
2778 GGGAACTGGAGAAACTATCATCCCTTAAAGAGGCCCTACAGCCACTG 2827  
2018 LeuThrGlnArgLeuProGlnLeuMetAlaProThrProProGlyLeu 2033  
2828 ATCAACAGAAAGATCCCTCAGTTATACAAAGCAGATATTGCCGTACCTG 2877  
2034 ...ArgAsnSerLeuAsnArgAlaSerPheArgLysAlaAspLeu 2047  
2878 CCACAGAGATCTCTCAGTCAAGTGAAGAGCTTCGCAAAATGATCTC 2923

seq\_name: gb\_pat:AX172978

seq\_documentation\_block:

LOCUS AX172978 3253 bp DNA

DEFINITION Sequence 105 from Patent WO0142297.

ACCESSION AX172978

VERSION AX172978.1 GI:14597957

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

Arbor Vita Corporation (US)

Clasp-3 transmembrane protein

Lu P., Garman, J.D. and Candia, A.F.

1 (bases 1 to 3253)

artificial sequence.

synthetic construct.

PAT 03-JUL-2001



```

1633 euleuGluaSpGluArgHisIleuProValGlyCysValSerPheGlnAsn 1649
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1586 TGCGGAGAGACCGGAATATCTTCTGCGGATGTGAACATTTTCACAAAT 1635
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1650 lIeSerSerAsnValIleuGluGluSerAlaIleSerAspAspIleuSe 1666
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1636 ATTTCATCTAATGTTTGAAGAATCGCGGTCCAGATGATGGTATC 1685
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1666 rProAspGluGluPheCysSerGlyLysHisPheThrGluLeuGlyL 1683
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1686 TCCGAGATGAAGAGATCTGCTCGAATAATCTTACTGAGTCAGAC 1735
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1683 euValGlyLeuLeuGluAlaIleAlaGlyTyrPheThrMetGlyGlyLeu 1699
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1736 TTGTGGGATTCGTGGAACAAGCAGCTGCTTCCTCTATGGCGTGATG 1785
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1700 TyrGluAlaValAsnGluValTyrLysAsnLeuIleProIleLeuGluAl 1716
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1786 TATGAGCAGATTAAAGATTACAAAGTACTTATTCCTATTCATGAGC 1835
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1716 aHisArgAspTyrLysLysLeuAlaIleValHisGlyLysLeuGluAla 1733
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1836 TAATCGGAGATGCAAGAACTATCCACAATTCATGCTGAACCTTCAGAG 1885
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1733 lArPheThrLysIleMetHisGlnSerSerGlyTyrPgluArgValPheGly 1749
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1886 CATTCACACAAATTTGTCATCAGAGTACTGCTGGGAGCGGATGTTGGC 1935
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1750 ThrTyrPheArGValGlyPheTyrGlyAlaHisPheGlyAspLeuAspGly 1766
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1936 ACCTATTTGCTGGTGGTTTATGAAACCAAGTTCGGGGATTTGGATGA 1985
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1766 uGlnGluPheValTyrLysGluProSerIleThrLysLeuAlaGluIleS 1783
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1986 ACAAGAAATTGTTTACAGAGAGCCTGCAATACCAAACTTCGACAGATAT 2035
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1783 eHisArGLeuGluGluPheTyrThrGluArPheGlyAspAspValVal 1799
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2036 CTCACAGATTGGAGGAGATTTCGAGAAAGATTGGAGAGATGGTGT 2085
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1800 GluIleIleLysAspSerTyrProValAspLysSerLysLeuAspSerGly 1816
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2086 GAAGTATTAACAAGACTTAATCTGTAGACAGAGTGAATTAATTAAGTCTAA 2135
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1816 nLysAlaTyrIleGlnIlePheTyrValGluProTyrPheAspThrTyrG 1833
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2136 CAAGGCATATATTCAGATTACTATGTGAGCCATCTTGAACACATATG 2185
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1833 lIleuLysAspArgValThrTyrPheAspArgAsnTyrGlyLeuArgThr 1849
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2186 AGATGAAGGACAGAAATCACCCTATTTGCAAAAAATTAACAATCTTCGTGA 2235
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1850 PheLeuPheCysThrProPheThrProAspGlyArgAlaHisGlyGluLe 1866
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2236 TTCATGTACTGTACACCTTTACTTAAAGTGGCGCTGCCCATGGGGAATC 2285
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1866 uProGluGlnHisLysArgLysThrLeuLeuSerThrAspHisAlaPheP 1883
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2286 TCATGAACAATTCMAAAGGAAGACATCTGACTACGCTTCATGGCCTTTC 2335
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1883 rOfTyrIleLysThrArgIleArgValCysHisArgGluGluThrValLeu 1899
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2336 CTTATATTAACAAGGTCATGTCATCTCATTAAGAAAGAGATCATCTTA 2385
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1900 ThrProValGluValAlaIleGluAspMetGlnLysLysThrArgGluLe 1916
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2386 ACACCAATTAAGAACTGTCTATGAGGACATCCAGAAAAAAGACACAGAGATT 2435
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1916 uAlaPheAlaThrGluGluAspProAspAlaLysMetLeuGlnMetV 1933
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2436 GGCATTTGCAACACATCAGAGATCCCGCACAGACCCCAAAATGCTTCAGATGG 2485

```

```

1933 allLeuGlnGlySerValGlyProThrValAsnGlnGlyProLeuGluVal 1949
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2486 TACTCCAGGATCTGTAGGACACACAGTGAATCAGGGGCTTTGGAAAGTT 2535
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1950 AlaGlnValPheLeuAlaGluIleProGluAspProLysLeuPheArgHis 1966
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2536 GCCCAGGTTTCTGTCTGAATTAACCTAGTGAACCCAAAGCTCTTCAGACA 2585
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1966 sHisAsnLysLeuArGLeuCysPheLysAspPheCysLysCysGluVal 1983
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2586 TCATATATAACCGCACCTCTGCTTAAAGATTTTACTTAAAGGCTGTGAAG 2635
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
1983 sPAlaLeuArGlyLysAsnLysAlaLeuIleGlyProAspGlnLysGluTyr 1999
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2636 ATGCCCTTAAGAAAAAATAAGACCTTAATTGGGCCGCTTCAAAAGAGATAT 2685
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2000 HisArgGluLeuGluArg. AsnTyrCysArgLeuArgGluAlaLeuGln. 2015
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2686 CAAAGGGAATTC...GGAAACTATCTTCGCTTAAAGAGGCCCTACAGC 2732
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2016 ProLeuLeuThrGlnArgLeuProGlnLeuMetAlaProThr.....Pr 2030
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2733 CCT...AGATCACAGAAAGTCCCTCAGTTATCCAAAGCCAGATATGCTTGT 2779
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2030 oProGlyLeuArGAsnSerLeuAsnArgAlaSerPheArGlyLysAlaAspL 2047
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2780 CCCCTGCCACAGAGATTCCTTCAGTCGAATGAGCTTTCGCAAAATGATC 2829
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2047 eu 2047
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
2830 TC 2831

```







OM of: US-09-736-968a-2 to: EST:\* out-format: pfs

Date: Oct 5, 2002 12:46 PM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framed-p2n.model -DEV=xlh  
-O=/cgn2\_1/USPRO.spool/US09736968/runat\_02102002.145334.1332/app-query.fasta\_1.2139  
-DB=EST -QPM=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.500  
-MINNAEXT=0.100 -LOPCPL=0.000 -LOPEXT=0.000 -GAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -EGAPOP=6.000  
-DELEX=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-LIST=45 -DOCLALIG=200 -MATRIX=tbls0m62 -TRANS=human40.cdi  
-THR\_SCORE=pc -THR\_MAX=100 -THR\_MIN=0  
-ALLEN=15 -MODE=LOCAL -OUTEXT=pfs -NORM=ext -HEAPSIZ=500  
-MINLEN=0 -MADE=LOCAL -USER=US09736968.@CGN1.1-6698  
-NCPD=6 -ICPD=3 -LONGLOG -DEV=US09736968 -WARN\_TIMEOUT=30  
-NO\_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-736-968a-2

Query length: 2047

Database: EST\*

Database sequences: 13736207

Database length: 1841457050

Search time (sec): 4657.170000

score\_list:

Sequence	Strd Orig	ZScore	EScore	len	Document
gb_hnc:AK016495	+ 4089.00	6097.36	0.0	2710	AK016495 Mus musculus adult ma
gb_hnc:BC020473	- 2212.00	3288.84	6.1e-174	1510	BC020473 Homo sapiens, clone 1
gb_hnc:AK010755	+ 2035.50	3024.09	3.4e-159	1507	AK010755 Mus musculus ES cells
gb_hnc:AK008416	+ 1819.50	2692.75	9.0e-141	2694	AK008416 Mus musculus adult ma
gb_hnc:BC018599	+ 1574.50	2334.25	9.0e-121	1311	BC018599 Mus musculus adult ma
gb_est2:BI770116	+ 1364.00	2024.71	1.6e-103	796	BI770116 603053332F1 NIH.MGC.12
gb_est2:BI782030	+ 1304.50	1933.54	1.9e-98	927	BI782030 603036864F1 NIH.MGC.11
gb_est2:BI830310	+ 1284.00	1903.12	7.2e-97	770	BI830310 603073042F1 NIH.MGC.11
gb_est2:BE531136	+ 1256.00	1862.55	1.7e-94	805	BE531136 601278315F1 NIH.MGC.39
gb_est2:BI854143	+ 1237.00	1863.49	7.0e-93	842	BI854143 603381452F1 NIH.CGAP.1
gb_est2:BI078974	+ 1213.00	1799.49	6.3e-91	777	BI078974 602873230F1 NCI.CGAP.1
gb_est2:BI181383	+ 1211.00	1795.62	9.0e-91	769	BI181383 UNL-P-FM-4E-08-0-UNL
gb_est2:BG678867	+ 1193.00	1767.26	3.4e-89	857	BG678867 60264767F1 NCI.CGAP.5
gb_est2:BI159195	+ 1178.00	1746.97	5.3e-88	778	BI159195 602922704F1 NIH.CGAP.1
gb_est2:AW742035	+ 1151.00	1706.93	7.9e-86	692	AW742035 up52604.Y1 Soares.mous
gb_est2:AJ397707	+ 1127.00	1656.17	1.0e-83	796	AJ397707 AJ397707 dkf426 Gal11u
gb_est2:BI822534	+ 1118.00	1656.50	5.1e-83	745	BI822534 603035242F1 NIH.MGC.11
gb_est2:BE857715	+ 1116.00	1655.50	6.0e-83	654	BE857715 7946603.X1 NCI.CGAP.1
gb_est2:AI653716	+ 1108.00	1647.98	2.9e-82	662	AI653716 wb36607.X1 NCI.CGAP.1
gb_hnc:AK018051	+ 1104.50	1625.81	2.6e-81	1712	AK018051 Mus musculus adult ma
gb_est2:BI181397	+ 1090.00	1613.37	1.3e-80	815	BI181397 603076025F1 NIH.MGC.11
gb_est2:BI110605	+ 1079.00	1598.32	8.6e-80	714	BI110605 60286211F1 NCI.CGAP.1
gb_est2:BE647357	+ 1065.50	1578.59	1.1e-78	696	BE647357 UI-M-BH1-amg-q-03-0-01
gb_est2:BG691673	+ 1064.50	1577.42	1.3e-78	678	BG691673 340944 BARC.5BOV Bos t
gb_est2:BG691673	+ 1057.00	1559.97	1.2e-77	1112	BI847359 AGENCOURT.646302 NIH
gb_est2:BI582283	+ 1045.00	1549.81	5.1e-77	644	AI582283 tg65109.X1 NCI.CGAP.1
gb_est2:AM516592	+ 1043.00	1546.45	6.8e-77	612	AM516592 xg01h03.X1 Soares.NHCG
gb_hnc:AK016777	+ 1043.00	1538.74	1.8e-76	1132	AK016777 Mus musculus adult ma
gb_est2:BM023626	+ 1011.00	1498.95	3.0e-74	588	BM023626 ieb2912.Y1 Melton Norm
gb_hnc:BC018075	+ 1008.00	1479.92	3.5e-73	1873	BC018075 Homo sapiens, clone 1
gb_hnc:BC018076	+ 1008.00	1479.92	3.5e-73	1873	BC018076 Homo sapiens, clone 1
gb_est2:BF980727	+ 993.50	1481.84	2.7e-73	938	BF980727 602333869F1 NIH.MGC.88
gb_est2:BG914864	+ 999.00	1479.55	3.6e-73	931	BG914864 602813690F1 NIH.CGAP.1
gb_est2:BG247108	+ 986.00	1455.68	7.8e-72	931	BG247108 602333976F1 NCI.CGAP.1
gb_est2:AM027227	+ 979.00	1450.44	1.5e-71	612	AM027227 wv66603.X1 Soares.thym
gb_est2:BM126124	+ 977.00	1448.02	2.1e-71	584	BM126124 if05h06.X1 Melton Norm
gb_est2:BG912304	+ 972.50	1438.97	8.6e-71	702	BG912304 602806839F1 NCI.CGAP.1
gb_est2:AW186286	+ 970.00	1437.42	6.1e-71	589	AW186286 xg62609.X1 NCI.CGAP.1
gb_est2:BF568737	+ 961.00	1432.84	4.1e-70	547	BF568737 602184333F1 NIH.MGC.42
gb_est2:BB659390	+ 961.00	1423.60	4.8e-70	604	BB659390 RIKEN full-16

seq\_name: gb\_hnc:AK016495

seq\_documentation\_block:

LOCUS

AK016495 2710 bp mRNA linear HTC 19-JAN-2002

DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched

library, clone:4931431C02:homolog to KIAA1395 PROTEIN (FRAGMENT),

full insert sequence.

ACCESSION AK016495.1 GI:12855259

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,

clone:lib:RIKEN full-length enriched mouse cDNA library

clone:4931431C02.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

1 (sites)

Carninci,P., Shihata,Y., Hayatsu,N., Sugahara,Y., Shihata,K.,

Itom,H., Konno,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

3 (sites)

Shihata,K., Itom,H., Mizawa,K., Nagakura,S., Sasaki,N., Carninci,P.,

Konno,H., Akiyama,U., Nishi,K., Kusunagi,T., Tashiro,H., Itom,H.,

Yamamoto,R., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,

Fujikawa,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M.,

Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,

Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A., and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system-184-format

sequencing pipeline with 384 multicapillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

4 (sites)

The RIKEN Genome Exploration Research Group Phase II Team and the

FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

5 (bases 1 to 2710)

Adachi,J., Atzawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,

Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bull,C.,

Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,

Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hirooka,T., Hori,F.,

Hume,D., Imotani,K., Ishii,Y., Itom,H., Iwama,M., Kasukawa,T.,

Kato,H., Kawai,T., Kojima,Y., Konno,H., Kouda,M., Koyama,S.,

Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,

Nunataki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,

Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,

Schimi,L., Shihata,K., Shihata,Y., Shingawa,A., Shiraki,T.,

Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,

Tanaka,T., Tejima,Y., Toyota,T., Yamamura,T., Yamane,K.,

Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M., and

Hayashizaki,Y.

Direct Submission

Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome

Exploration Research Group, RIKEN Genomic Sciences Center (GSC),



```

|||||
952 TACTGAAGACATTGTGAGAAAGTCCCTCAAGACATCTGACCTGACCGAG 1001
1537 LUGLASPmetGlyLeuArgSerThrPheAlaGluInValGlnAsp 1553
1002 AGGAGGACATAGGGCTGAGGAGACAGCAGCTCGCTGAGCAGGTCAGAGAC 1051
1554 LeuMetPheAsnLeuHisMetIleLeuThrAspThrValLysMetLysG1 1570
1052 CTCATGTTCAACCTGCACATGATCTGACAGACACGCTGAAGATGAAAGGA 1101
1570 uHisGlnGluAspProGluMetLeuIleAspLeuMetTyrArgIleAla 1587
1102 ACACGAGGAGGAGACCTGAGATGCTGATGAGACTCATGTACAGAAATTGCTC 1151
1587 rGlyTyrGlnGlySerProAspLeuArgLeuThrTyrPleuGlnAsnMet 1603
1152 GGGGATACCAAGGCTCTCCAGACCTGGGCTGACATGGCTGCAGAAACATG 1201
1604 AlaGlyLysHisAlaGluLeuGlyAsnHisAlaGluAlaGlnCysMe 1620
1202 GGTGCAAAATGCAAGCTGGGCAATCATGCAAGGCTGCCAGTGCAT 1251
1620 tValHisAlaAlaAlaLeuValAlaGluTyrIleuAlaLeuGluAspG 1637
1252 GGTCCAGCGCGCTGCTGGTGGCCAAATACCTTGGCTCTCTCGAGAGACA 1301
1637 LArgHisLeuProValGlyCysValSerPheGlnAsnIleSerSerAsn 1653
1302 GTCCGACACCTGCTGGGCTGTGTCTTCAGAAATGTCATCAACAC 1351
1654 ValLeuGluGluSerAlaIleSerAspAlaIleLeuSerProAspGlu 1670
1352 GTGCTGAGAGAGTGCACATCTGTATGACATCTGTCCGAGATGAGGA 1401
1670 uGlyPheCysSerGlyLysHisPheThrGluLeuGlyLeuValGlyLeu 1687
1402 GGGCTTCTCTCTAGSAAACTTCACAGAACTGGTCTGGTGGGGCTGC 1451
1687 euGluGlnAlaAlaGlyTyrPheThrMetGlyGlyLeuTyrGluAlaVal 1703
1452 TGGAGAGGCGGCGGCTACTTCACCATGGTGTGTATGAAACGGG 1501
1704 AsnGluValTyrLysAsnLeuIleProIleLeuGluAlaHisArgAsp 1720
1502 AACGAGGTACAAACCTTATCCCATCTTGAAGCCACAGAGACTA 1551
1720 rLysLysLeuAlaAlaValHisGlyLysLeuGlnGluAlaPheThrLys 1737
1552 CAAGAAGCTGGCTGGTGCAGGGAACCTGCAGAGGCTTCACCAAGA 1601
1737 LeuMetHisGlnSerSerGlyTyrGluArgValPheGlyTyrThrPheArg 1753
1602 TTATGACACAGAGCTCTGGCTGGAGCGGTGTATTGGGACATATTCCGA 1651
1754 ValGlyPheThrGlyAlaHisPheGlyLysPleuAspGluGlnGluPhe 1770
1652 GTGGGCTTATGACACAGATTGTGACCTGGATGAACAAGACTTTGT 1701
1770 LArgLysGluProSerIleThrLysLeuAlaGluIleSerHisArgLeu 1787
1702 GTACAAAGGACCGTCAATCAGCAACCTTGCAGAGATCTCACACCGGCTGG 1751
1787 LUGLUPhetTyrThrGluArgPheGlyAspAspValAlaGluIleLys 1803
1752 AGGAGTTCATACGGAAGGTTCCGGGATGATGTGTAGAGATCATCAAA 1801
1804 AspSerTyrProValAspLysSerLysLeuAspSerGlnLysAlaTyr 1820
1802 GATTTCATACCCGCTGACAAAGTCCAAAGCTGGACCCAGAAAGGCGTTAAT 1851
1820 eGluIleThrTyrValGluProTyrPheAspThrTyrGluLeuLysAsp 1837
|||||

```

```

1852 ACAGTAAACCTATGTGGAGCCACATTTCCGACATTTATGAGCTCAAGATC 1901
1837 rGValThrTyrPheAspArgAsnTyrGlyLeuArgThrPheLeuPheCys 1853
1902 GGGTACCTACTTCGATTCGGAACATATGGCTGCGGGCTTCTTCTTCTGC 1951
1854 ThrProPheThrProAspGlyArgAlaHisGlyGluLeuProGluGlnH 1870
1952 ACACCTTCACACACAGATGAGAGCTGGCAGCAGAGAGTGGCCGAACAGA 2001
1870 sLysArgLysThrLeuLeuSerThrAspHisAlaPheProGlyTyrIle 1887
2002 CAAGCGAAGAGCGCTGAGACGAGAGATGCTTCCCTTCATATCAGA 2051
1887 hrArgIleArgValCysHisArgGluGluThrValLeuThrProValGlu 1903
2052 CACGATCCGAGTGTGCCACCGTGAAGAGACAGTGTGACACCGAGTGAG 2101
1904 ValAlaIleGluAspMetGlnLysLysThrArgGluLeuAlaPheAla 1920
2102 GTGGCATTTGAGGACATGCAGAAAGACCCGGAGCTGGGCTTGGCCAC 2151
1920 rGluGlnAspProProAspAlaLysMetLeuGlnMetValLeuGlnLys 1937
2152 CGAGCAGACCTTCAGATGCAAGATGCTGAGATGATGCTTCAGAGGTT 2201
1937 erValGlyProThrValAsnGlnGlyProLeuGluValAlaGlnValPhe 1953
2202 CTGGGGAGCCACAGTGTGAACAGGCTCCCTTGGAAAGTGGCCAGCTGTT 2251
1954 LeuAlaGluIleProGluAspProLysLeuPheArgHisHisAsnLys 1970
2252 TTGTGAGAGATCCAGAAAGATCCCAAGCTTCCGACATCAACAACACT 2301
1970 uArgLeuCysPheLysAspPheCysLysLysCysGluAspAlaLeuArg 1987
2302 CCGGCTGCTTCAAGATTTTCTCAAAAGTGGAGAGTGCACGTGGA 2351
1987 yAsnLysAlaLeuIleGlyProAspGlnLysGluTyrHisArgGluLeu 2003
2352 AGAACAGAGCCCTGATTTGGCCAGACAGAAAGATACCAACCGGAGCTG 2401
2004 GLuArgAsnTyrCysArgLeuArgGluAlaLeuGlnProLeuLeuThr 2020
2402 GAGGCTCACTATACCGCTGCGGAGAGCTGTGCAGCCTGCTTACCCA 2451
2020 nArgLeuProGlnLeuMetAlaProThrProProGlyLeuArgAsnSer 2037
2452 ACGTCTGCCCAAGCTGCTGGCACCAGATTCACCAAGCCTCAGAGGCTCA 2501
2037 euAsnArgAlaSerPheArgLysAlaAspLeu 2047
2502 TGAACAGATCAAGTTTCAGGAAGGCTGACCTC 2533
seq_name: gb_hc:BC020473
seq_documentation_block:
LOCUS BC020473 1510 bp mRNA linear HTC 03-JAN-2002
DEFINITION Homo sapiens, clone IMAGE:3853958, mRNA.
ACCESSION BC020473
VERSION BC020473.1 GI:18042973
KEYWORDS HTC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1510)
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

```

REMARK  
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [villation@bcm.tmc.edu](mailto:villation@bcm.tmc.edu)  
Villation, D.K., Luna, R.A., Hale, S.M., Huiyk, S., Lu, X., Garcia,  
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,  
Muzny, D.M., Gunaratne, P., Yeom, V., Kowis, C., Martin, R.,  
Lawrence, S., Richards, S., Gibbs, R.A.

## FEATURES

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAC Plate: 13 Row: b Column: 12  
This clone has the following problem: no polyA-tail.  
Location/Qualifiers

## SOURCE

1. 1510  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3853958"  
/tissue\_type="Colon, adenocarcinoma"  
/clone\_lib="NHL MGC\_65"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"

BASE COUNT 287 a 431 c 456 g 336 t  
ORIGIN

## alignment\_scores:

Quality: 2212.00 Length: 425  
Ratio: 5.229 Gaps: 0  
Percent Similarity: 99.529 Percent Identity: 99.294

## alignment\_block:

US-09-736-968A-2 x BC020473/rev ..

Align seg 1/1 to reverse of: BC020473 from: 1 to: 1510

```
1623 AAlaAlaLeuValAlaGluTyrLeuAlaLeuLeuGluAspGlnArgHt 1639
|||||
1510 GCGCGCGCGCTCGTGGCTGATACCTCGCCCTGCTGAGAGACACCGCA 1461
|||||
1639 GTGCTCGGGAGGACATTCAGTGGGCTGGTGGGCTTCTGGAACA 1311
|||||
1460 CCTGCCCGTGGGCTGCTTCTCCAGAACATCTCATCATGTGCTAG 1411
|||||
1656 TugLusSerAlaIleSerAspAlaLeuSerProAspGluGlyPhe 1672
|||||
1410 AGGAGTCCGCATCTCCGAGACATCTGCGCCGACGAGAGGGCTTC 1361
|||||
1673 CysSerGlyLysHisPheThrGluLeuGlyLeuValGlyLeuLeuGlu 1689
|||||
1360 TGCTCCGGGAGGACATTCAGTGGGCTGGTGGGCTTCTGGAACA 1311
|||||
1689 nAlaAlaGlyTyrPheThrMetGlyGlyLeuTyrGluAlaValAsnGlu 1706
|||||
1310 GGCAGCGGCTACTTACCATGGGCGGCTCTACGAGCGGTGATGAGG 1261
|||||
1706 aTyrLysAsnLeuIleProIleLeuGluAlaHisArgAspTyrLys 1722
|||||
1260 TCTACAGAAACATCCCATCTCGAAGCCACCGTACATACAAAGAG 1211
|||||
1723 LeuAlaAlaValHisGlyLysLeuGluGluAlaPheThrLysIleMet 1739
|||||
1210 CTGGCGCGGTGCACGCAAACTGCGAGGAGGCTTCCACCAAGATATGCA 1161
|||||
1739 sGlnSerSerGlyTyrGluArgValPheGlyThrTyrPheArgValGly 1756
|||||
```

```
1160 CCAGAGTTCGGGCTGGGAGCGCGTGTTCGGGAGCTATTTCGGCGTGGGCT 1111
|||||
1756 heTyrGlyAlaHisPheGlyAspLeuAspGluGluGluPheValTyrLys 1772
|||||
1110 TCTACGCGCGGCTACTTGGTGGTGGCTGATGAGCAGGAGTTGTGTACAG 1061
|||||
1773 GluProSerIleThrLysLeuAlaGluIleSerHisArgLeuGluGlu 1789
|||||
1060 GAGCCATCATCATCAGAGCTGGGAGAGATCTCACACCGGCTGGAGGAGTT 1011
|||||
1789 eTyrThrGluArgPheGlyLysAspValValGluIleIleLysAspSer 1806
|||||
1010 CTACACGAGAGATTGGCGACGCGTGTGAGATTATCAAGAGCTCTA 961
|||||
1806 YrProValAspLysSerLysLeuAspSerGlnLysAlaTyrIleGlnIle 1822
|||||
960 ACCCTGTGGACAACTCCAACTTGACTACAAAAGGCTCATTCAGATC 911
|||||
1823 ThrTyrValGluProTyrPheAspThrTyrGluLeuLysAspArgVal 1839
|||||
910 ACGTATGTGGAAACGCTACTTGTATACCTACAGACTCAAGAGCGGGTAC 861
|||||
1839 rTyrPheAspArgAsnTyrGlyLeuArgThrPheLeuPheCysThrProp 1856
|||||
860 CTACTTTGACCGCAACTATGAGGCTTCCACATCTCTTGTGACGCGCT 811
|||||
1856 heThrProAspGlyArgAlaHisGlyLysLeuLeuProGluGlnHisLysArg 1872
|||||
810 TCACGCGCGATGGGCGGCGCACGCGGAGAGCTGCCGCAACAAAGCGT 761
|||||
1873 LysThrLeuLeuSerThrAspHisAlaPheProTyrIleLysThrArg 1889
|||||
760 AAGAGCTGTCTCAGCACCGACCGACCTTCCCTCATCAAGACTGCGCAT 711
|||||
1889 eArgValCysHisArgGluGluThrValLeuThrProValGluValAla 1906
|||||
710 CCGTGTGTCCACCGGAGAGAGAGAGGCTGCTGACGCCGCTGAGAGTGCCA 661
|||||
1906 IeGluAspMetGlnLysLysThrArgGluLeuAlaPheAlaThrGluGln 1922
|||||
660 TCGAGGACATGCAGAGAGAGACGAGAGCTTGCCCTTCCACCGAGCAG 611
|||||
1923 AspProProAspAlaLysMetLeuGlnMetValLeuGlnIleGlySerVal 1939
|||||
610 GACCACACAGATGCTAAGATCTACAGATGGTCTTACAGGCTCTGTAGG 561
|||||
1939 YrProThrValAsnGlnGlyProLeuGluValAlaGlnValPheLeuAla 1956
|||||
560 GCCACCGTGAACACGAGGTCCCTCGAGAGTGGCCACAGTGTTTTATCAG 511
|||||
1956 LuIleProGluAspProLysLeuPheArgHisHisAsnLysLeuArgLeu 1972
|||||
510 AGATCCGGAAGACCCCAAGCTTTCGCGCATCCAACTTCGGGCTC 461
|||||
1973 CysPheLysAspPheCysLysLysCysGluAspAlaLeuArgLysAsn 1989
|||||
460 TGCTTCAAGGACTTCTGAAAGAAATGAGAGATGCGTGGCGAAATAA 411
|||||
1989 sAlaLeuIleGlyProAspGlnLysGluTyrHisArgGluLeuGluArg 2006
|||||
410 GGCCTGATTGGGCGGAGCAGAAAGAGTACCAACGAGAGCTGAGCGCA 361
|||||
2006 sNTrpCysArgLeuArgGluAlaLeuGlnProLeuLeuThrGlnArgLeu 2022
|||||
360 ACTACTGCGGCTCGGAGAGCTCTGAGGCCCTTACCCACAGCGGCTG 311
|||||
2023 ProGlnLeuMetAlaProThrProProGlyLeuArgAsnSerLeuAsn 2039
|||||
310 CCCCAGCTGATGACACCCACCCGCGGCTTCAGAACTCTTGAACAG 261
|||||
2039 gAlaSerPheArgLysAlaAspLeu 2047
|||||
260 AGCAAGTTCCGAAGGACAGACCTC 236
```



```

253 TCTACATGTCAGCTGCCAGCCCGGCCAGGAAACAGCCTTGAGAGCGCT 302
654 ValGlyPheThrTrpIleProLeuLeuGlnHisGlyArgLeuArgTrpG 670
303 GTGGGCTTCACCTGGATCCCTCTGTACACATGCGCGCTGAGAGACTGG 352
670 yProPheValLeuProValSerValAspGlnProProSerSerTyrSer 687
353 TCCCTTGCCCTGGCCGCTGTCGTGAGACACCTCCACCCAGCTACTCGG 402
687 aLeuThrProAspValAlaLeuProGlyMetArgTrpValAspGlyHis 703
403 TCCGTACCCAGATGTAGCGCTGCTGGCATGCGCTGGTGGAGCGGCAC 452
704 LysGlyValPheSerValGlyLeuThrAlaValSerSerValHisProG 720
453 AAGGCGGTTCAGTGTGAGCTCACCGCCGTGTCTGTCGTCAGCCGCA 502
720 nasProTyrLeuAspLysPhePheThrLeuValHisValLeuGlnG 737
503 GACACCCCACTTGGATAGTTCTTCACGCTGTGATCATTTCTAGAGAG 552
737 yAlaPheProPheArgLeuLysAspThrValLeuSerGlyLysVal 753
553 GGATCTTCCGTTCCGACTCAAGGAGACAGTGTGAGCTAAGGCACTATG 602
754 GluGlnGluLeuArgAlaSerLeuAlaAlaLeuArgLeuAlaSerProG 770
603 GAGCGAGAGTTGCCGAGCAGCTGGACCCCTGCGCTTGACGCCGCA 652
770 uProLeuValAlaPheSerHisHisValLeuAspLysLeuValArgLeu 787
653 GCCCTTAGAGCCTTTCACACCTGCTGCTGACAGAGTTTCCGCTGG 702
787 aIleArgProProIleLeuSerGlyGlnIleValAsnLeuGlyArgGly 803
703 TTGTGCGGCGACCATCATTTGTGGCAGATGGTGAACCTGGTGAGAGG 752
804 aAlaPheGluAlaMetAlaHisValAlaSerLeuValHisArgSerLeuG 820
753 GCGTTTGAAGCTATGCTATGTACACAGCCTGTGCACCGAACCTTGA 802
820 uAlaIleAspAlaArgGlyHisCysProGlnLeuAlaIleArgValH 837
803 GGCTGCGCAAGACTCCCGTGGCCACTGGCCACTGTGGCTCCCTATGTC 852
837 aSTyrAlaPheArgLeuProGlyThrGluProSerLeuProAspGlyAla 853
853 ACTACGCCCTTTCGCTCCTCGTGTGACTCCTGACCTCCAGCCAGGCA 902
854 ProProValThrValGlnAlaAlaThrLeuAlaArgGlySerGlyArgPr 870
903 CCTCCGACAGCTGTGAGGCTCCCACTGGCCCTGGCTGTGTCGCC 952
870 oAlaSerLeuTyrLeuAlaArgSerLysSerIleSerSerSerAsnPro 887
953 CGGACACCTGTACTGGCAGCGGTCTAAGAGCATCAGACAGACACCTG 1002
887 spleuAlaValAlaProGlySerValAspAspGlyValSerArgIleLeu 903
1003 ACTTGCTGTGGTCCCTGCTGTGTGAGAGATGATGTCGCGCATCTTA 1052
904 Alaser..... 905
1053 GCCAGCAGGCGCATGACCGCAACTGTAGCCGGGCTCTTACTTACTGA 1102
906 .....L 906
1103 GGCGTCATCTCGGCCCTGCAGCCCTCAGTCGAGACACACTGTGCAGA 1152
906 yLeuLeuHisGluGluLeuAlaLeuGlnTrpValValSerSerSerAla 922
1153 AGCTGCTGACGAGAGCTGGCTGTGCACTGGTGGTGAGCGGCGAGTGGC 1202

```

```

923 ValArgGluAlaIleLeuGlnHisAlaIlePhePhePheGlnLeuMetVa 939
1203 GTTCGTGAGCTAGTCTCTGCAGCATGCGCTGTTCTTCTTCACCTATG 1252
939 LysSerMetAlaLeuHisLeuLeuLeuGlnGlyAlaArgLeuAspThrPro 956
1253 AAAAAGCATGAGACTTCATCTGCTGTGGCCAGCGAGTGCAGACTGCC 1302
956 rGlyLeuAlaArgPheProGlyArgPheLeuAspAspIleThrAlaLeuVal 972
1303 GCAACGTGGCGCTTCCCTGGGCGCTTCTGTGATGACATTCGCCCTGGTG 1352
973 GlySerValGlyLeuGlnValIleThrArgValHisLys 985
1353 GCTTCGCTGGCGCTGGAAGTCATCACCAGAGTCCATTAAG 1391

seq_name: gb_hlc:AK004816
seq_documentation_block:
LOCUS AK004816 2694 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus adult male lung cDNA, RIKEN full-length enriched
library, clone:1200017A24:homolog to BA165P24.1.1 (NOVEL PROTEIN
SIMILAR TO DROSOPHILA CG6630 AND CG11376, KIAA1058, RAT TRG
(SISOFORM 1)) (FRAGMENT), full insert sequence.
ACCESSION AK004816
VERSION AK004816.1 GI:12836282
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male lung cDNA to mRNA,
clone:1200017A24.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (sites)
AUTHORS Carninci,P., and Hayashizaki,Y
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
2 (sites)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
3 (sites)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kasaiwa,K.,
Fujitake,S., Inoue,K., Togawa,Y., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Yoneda,Y., Ishikawa,T., Kira,A., and Hayashizaki,Y.
TITLE Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A., and Hayashizaki,Y.
Riken integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
4 (sites)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
PUBMED 11076861
5 (bases 1 to 2694)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Arikawa,T., Balderelli,R., Bono,H., Brownstein,M., Bulic,C.,
Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hirooka,T., Hori,F.,
Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,

```





```

1818 1aTyrTleGlnIleThrTyValGluProTyrPheAspThrTyGluLeu 1834
      |||
      |||
      |||
845 CCTACATTGATCATCTTTGTGAGCCTTACTTGTGATGAAATATGAGATG 894
      |||
      |||
      |||
1835 LysAspArgValThrTyrPheAspArgAsnTyrGlyLeuArgThrPheLe 1851
      |||
      |||
      |||
895 AAAGACCGGGTGACCTTACTTGAGAAAGAAATTCAACTCCGAGGTTTCAT 944
      |||
      |||
      |||
1851 uphEysThrProPheThrProAspGlyArgAlaHisGlyGluLeuProG 1868
      |||
      |||
      |||
945 GTACACCCACCCCTTCACCCCTGAGAGGAGACCCCGGGGCGAGCTTCATG 994
      |||
      |||
      |||
1868 LueGlnHisLysArgLysThrLeuLeuSerThrAspHisAlaPheProTyr 1884
      |||
      |||
      |||
995 AGCAACACCCGAGAAACACCGTCTCACCACCATCAGCGCTTCCTCCCTAC 1044
      |||
      |||
      |||
1885 IleLysThrArgGlyLeuArgValCysHisArgGluGluThrValLeuThrPr 1901
      |||
      |||
      |||
1045 ATCAAGACCGAGATCCGAGTCAGCCAGAAAGAGAGTTCGTTTGACTCC 1094
      |||
      |||
      |||
1901 oValGluValAlaIleGluAspMetGlnLysLysThrArgGluLeuAlaP 1918
      |||
      |||
      |||
1095 GATTGAGGTTCCCATGTGAAGATATGAGAAAGAACACCCCTGCATTAACCG 1144
      |||
      |||
      |||
1918 heaLarThrGluGlnAspProProAspAlaLysMetLeuGlnMetValLeu 1934
      |||
      |||
      |||
1145 TGCGCCTACACAGAGAGCCCTCATCAAAAGATGTCGAAATGCTACTG 1194
      |||
      |||
      |||
1935 GlnGlySerValGlyProThrValAsnGlnGlyProLeuGluValAlaG 1951
      |||
      |||
      |||
1195 CAGGGCTCTGAGAGCCACCTGTAAATACAGGACCTGGAGGGGCCCA 1244
      |||
      |||
      |||
1951 nValPheLeuAlaGluIleProGluAspProLysLeuPheArgHisHis 1968
      |||
      |||
      |||
1245 AGTGTTCTTGCTGAAATTCACAGCTGACCCAAAGCTCTACCGAATCACA 1294
      |||
      |||
      |||
1968 snLysLeuArgLeuCysPheLysAspPheCysLysLysCysGluAspAla 1984
      |||
      |||
      |||
1295 ACAAGCTGAGGTGTGCTTCAAGAGATTCAATATGATCGATCGGAGAGGCC 1344
      |||
      |||
      |||
1985 LeuArgLysAsnLysAlaLeuIleGlyProAspGlnLysGluTyrHisSar 2001
      |||
      |||
      |||
1345 GTGGAGAAAGAACAGCGCATCTATCACCAGAGAGCGGGGATGCCAGCA 1394
      |||
      |||
      |||
2001 gGluLeuGlnArgAsnTyrCysArgLeuArgGluAlaLeuGlnProLeu 2018
      |||
      |||
      |||
1395 GGAGCTGAGAGAACTACAAACAGCTGAGAGACAGCTCAGGCCCATGA 1444
      |||
      |||
      |||
2018 eutHrGlnArgLeuProGlnLeuMetAlaPro.....ThrProPro 2031
      |||
      |||
      |||
1445 TTGAGCGGAAATCCAGAGCTCTACAAAGCCCATATTCTCAGAGTTGACAGT 1494
      |||
      |||
      |||
2032 GlyLeuArgAsnSerLeuAsnArgAlaSerPheArgLysAlaasp 2046
      |||
      |||
      |||
1495 CAGAAGAGGAGCTCTTCCACAGATCTAGTTTCAGAAATGTGA 1539
      |||
      |||
      |||
seq_name: gb_hc:BC018599
seq_documentation_block:
LOCUS BC018599 1311 bp mRNA linear HTC 03-DEC-2001
DEFINITION Mus musculus, clone IMAGE:3596745, mRNA.
ACCESSION BC018599
VERSION BC018599.1 GI:17235789
KEYWORDS HTC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer

```

```

REMARK
COMMENT
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.ncl.nih.gov
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villaloboscm.tmc.edu.
Villalobos, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 16 Row: e Column: 2
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis
This clone has the following problem: no 5' EST match.
FEATURES
location/Qualifiers
1..1311
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3596745"
/russus_type="Mammary tumor, C3(1)-Tag model. Infiltrating
ductal carcinoma, 5 month old virgin mouse."
/clone_id="NCI_CGAP_Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
BASE COUNT 363 a 309 c 333 g 306 t
ORIGIN
alignment_scores:
Quality: 1574.50 Length: 440
Ratio: 4.048 Gaps: 5
Percent Similarity: 88.409 Percent Identity: 71.364
alignment_block:
US-09-736-968a-2 x BC018599 ..
Align seg 1/1 to: BC018599 from: 1 to: 1311
1197 AspSerAspThrGlnGluGlyLysPheAlaGlyThrIleAsnProse 1213
      |||
      |||
      |||
16 GACTATGACAGTGAAGT.....GGGACATGATTAAGCCAGAC 53
      |||
      |||
      |||
1213 rValAlaMetAlaIleAlaGlyGlyProLeu.....AlaProG 1226
      |||
      |||
      |||
54 GGTTCGATGCGCATTTGCGAGAACATCCGTCCTCAGCTGACAAACCTG 103
      |||
      |||
      |||
1226 LysSerArgAlaSerIleSerGlnGlyProProThrAlaSerArgAlaGly 1242
      |||
      |||
      |||
104 CGAGTTTCTCCTCAGCTCA.....ACGAGTGGCGGCAGCAC 141
      |||
      |||
      |||
1243 CysAlaLeuSerAlaGluSerSerArgThrLeuLeuAlaCysValLeuTr 1259
      |||
      |||
      |||
142 ACCACCTTCTCAGACAGAGTCAAGAGAGCCCTTTGATCTGCTGCTTGG 191
      |||
      |||
      |||
1259 pValLeuLysAsnThrGluProAlaLeuLeuGlnArgTrpAlaThrAspL 1276
      |||
      |||
      |||
192 GGTTCGACAGATGACAGATGAACGTTCTGCAGAAATGGTTTACAGATC 241
      |||
      |||
      |||
1276 eutHrLeuProGlnGlnGlyArgLeuLeuAspLeuLeuTyrLeuCysLeu 1292
      |||
      |||
      |||
242 TCTCAGTCTCCAGCTGAACCGTGTGTAAGATCTGTTTACCTGTGTGTA 291
      |||
      |||
      |||
1293 AlaAlaPheGluTyrLysGlyLysAlaPheGlnAlaGlyIleAsnSerLe 1309

```

```

1309 uhrPhetLysSerLeuAspMetLysAlaArgLeuGluValaIleL 1326
1342 AACCTTTAGAAATCAAAAGACATGAGAGCCAAAGCTTGAGAGAGCCATTTC 391
1326 euGIYThrIleGlyAlaArgGluMetValaArgSerArg..... 1340
1392 TGGGGGCACTGGTGCAGAGAAATGGTACGGCGGAGCGGAGGCGAG 441
1341 ...GluArgSerPro.....PheGlyAsnProGluAsnValaArg 1352
1342 CTCGAGAGAGAGCCCATCTGGAAGCCCTTGGGAGCCAAAGAAACCTGCG 491
1352 gTTPArgLysSerValaThrHisTrpLysGlnThrSerAspArgValaAspL 1369
1352 ..... 1369
492 GTGGAGAAAGACATGACTGCTGGCGTGCAGAACTCAGATA 541
1369 ySThrLysAspGluMetGluHisGluAlaLeuValaGluGlyAsnLeuAla 1385
1369 ..... 1385
542 AGTCAAGACACAGAGATGAGACAGAGCACTGATGATGAGAAACCTGGCT 591
1386 ThrGluAlaSerLeuValaValaLeuAspThrLeuGluIleIleValaGlnTh 1402
1386 ..... 1402
592 ACTGAAGAAACCTCTCATCTTATAGACAGCTAGAGATCATGCTCCAGAC 641
1402 rValMetLeuSerGluAlaArgGluSerValaLeuGlyAlaValaLeuLysV 1419
1419 ..... 1419
642 TCGTGTCTGAACGGAATCCAAAGAGAGTTCGGTGGTGGTAAAG 691
1419 aValLeuLysSerLeuGlySerAlaGlnSerAlaLeuPheLeuGlnHis 1435
1435 ..... 1435
692 TGCCTGTACAGACATGGCTCCAAACCAAGTGCAGTCTATCTGCAGCAC 741
1436 GlyLeuAlaThrGlnArgAlaLeuValaSerLysPheProGluLeuLeuPhe 1452
1452 ..... 1452
742 TGGTTCGCCACGAGAGAGCCCTGGTCTCAAAAGTTTCCGTGACTCTGTT 791
1452 eGIuGluAspThrGluLeuCysAlaAspLeuCysLeuArgLeuLeuArgH 1469
1469 ..... 1469
792 TGAAGAGAGACAGACAGACTGCGGATCTGCCCTCCGCTTCGCCGTC 841
1469 iScyGlySerArgLysSerThrIleArgThrHisAlaSerAlaSerLeu 1485
1485 ..... 1485
842 ATTTGAGTACAGACATCAGTACACATCCGGTGCAGCTAGTCCCTCCTT 891
1486 TyrLeuLeuMetArgGlnAsnPheGluIleGlyHisAsnPheAlaArgVala 1502
1502 ..... 1502
892 TACCTCTCATGAGCGAGAACTTGGATCGGGAATACTTTGCCAGAGT 941
1502 LysMetGlnValaThrMetSerLeuSerSerLeuValaGlyThrThrGlnAla 1519
1519 ..... 1519
942 GAAATGCAAGTGACGATGCTCACTGCTTGGTGGGTACGTCAGCA 991
1519 snPheSerGluGluHisLeuArgArgSerLeuLysThrIleLeuThrTyr 1535
1535 ..... 1535
992 ATTTATATAGAAATCTTAAGACGCTCTTAAACCTAATTTTGAACATAT 1041
1536 AlaGluGluAspMetGlyLeuArgAspSerThrPheAlaGluGlnValaGln 1552
1552 ..... 1552
1042 GCGAAGAAAGATCTAGAATTGAGGGAACACATTTCCGTATCAGGTCCA 1091
1552 nAspLeuMetPheAsnLeuHisMetIleLeuThrAspThrValaLysMetL 1569
1569 ..... 1569
1092 AGATTGTGCTTCATATCTCATATGATCTTCTCTGACACTGTTAAGATGA 1141
1569 ySGluHisGlnGluAspProGluMetLeuIleAspLeuMetLysArgLys 1585
1585 ..... 1585
1142 AGGAGCATAGAGAAAGCCAGAAATGCTGATTCATGTGACAGAAATC 1191
1586 AlaArgGluTyrGlnGlySerProAspLeuArgLeuThrTrpLeuGlnAs 1602
1602 ..... 1602

```

```

1192 GCCAAGGGCTACGAGAGCTCTCCAGAGACTGGCGCTGACTGTTGCAGAA 1241
1602 nMetAlaGlyLysHisAlaGluLeuGluGlyAsnHisAlaGluAlaGlnC 1619
1242 CATGCTGGCGAAGCATCTGTAACGAGAACATTCATGCGAAGCTGTCTCAGT 1291
1619 ySMetValHisAlaAlaAla 1625
1292 GCGTGTCTCACTCCGAGCC 1311
seq_name: gb_est2:B1770146

```

```

seq documentation block:
LOCUS      B1770146              796 bp      mRNA      linear      EST 25-SEP-2001
DEFINITION 603053352F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5203063 5',
            mRNA sequence.
ACCESSION  B1770146
VERSION    B1770146.1 GI:15761724
KEYWORDS   EST.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 796)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cga@bbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            plate: LLNL1508 row: p column: 08
            high quality sequence stop: 793.
            Location/Qualifiers

```

```

FEATURES
    source
        1..796
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="IMAGE:5203063"
            /clone_lib="NIH_MGC_122"
            /lab_host="DH10B"
            /note="Organ: Pooled lung and spleen; Vector: pCMV-Sport6;
            site_1: NotI; Site_2: EcoRV (destroyed); RNA source
            anonymous pool of 24 week female lung, 16 week female
            spleen, and 20-22 week male spleens. Library is oligo-dT
            primed and directionally cloned (EcoRV site is destroyed
            upon cloning). Average insert size 1.4 kb, insert size
            range 1-3 kb. Library is normalized and enriched for
            full-length clones and was constructed by C. Gruber
            (Invitrogen). Research Genetics tracking code 026. Note:
            this is a NIH_MGC Library."

```

```

BASE COUNT      188 a      228 c      231 g      148 t      1 others
ORIGIN

```

```

alignment_scores:
    Quality: 1364.00      Length: 264
    Ratio: 5.226          Gaps: 2
    Percent Similarity: 98.864      Percent Identity: 98.864

```

```

alignment_block:
US-09-736-968A-2 x B1770146 ..

```

```

Align seg 1/1 to: B1770146 from: 1 to: 796

1755 glyPheTyrGlyAlaHisPheGlyAspLeuAspGluGlnGluPheValaTyr 1771
|||||
1 GCGCTTCTACGGCGCCCACTTGGTACCTGGATGAGCAGAGAGTTTGTTGTA 50
|||||

```

```

1771 rlysgluProSerIleThrLysLeuAlaGluIleSerHisArgLeuGlu 1788
|||||
51 CAAGGAGCCATCGATCAGACAGTGCAGAGATCTCACACCGCTGGAGG 100
|||||
1788 lueherythrGluArgPheGlyAspAspValValGluIleLeuLysAsp 1804
|||||
101 AGTTCTACACGGAGAGATTGGCGAGACGCTGTTGAGATTATCAAAAGAC 150
|||||
1805 SerTyrProValAspLysSerLysLeuAspSerGlnLysAlaTyrIleG1 1821
|||
151 TCFAACCCCTGTGACAAAGTCCAAGCTTGACTCACAAAAGCCCTACATCCA 200
|||||
1821 nilethrrValGluProTyrPheAspThrTyrGluLeuLysAspArgV 1838
|||||
201 GATCAGCTATGTGGAAACCTTACTTGATACCTACAGCTCAAGACCGGG 250
|||||
1838 alThrTyrPheAspArgAsnTyrGlyLeuAlaArgThrPheLeuPheCysThr 1854
|||||
251 TGACCTACTTGTGACCGCAACTATGGGCTTCGACATCTCTGTTCTGCACG 300
|||||
1855 ProPheThrProAspGlyArgAlaHisGlyGluLeuProGluGlnHisLy 1871
|||||
301 CGCTTCACCGCCGGATGGCGCGACACGCGGAGCTGCCGACCAACACAA 350
|||||
1871 sarGlyThrLeuLeuSerThrAspHisAlaPheProTyrIleLysThrA 1888
|||||
351 GCGTAGAGAGCTGCTCAGCACCGACCGCTCCCTCCCTACATCAAGACTTC 400
|||||
1888 rglleArgValCysHisArgGluGluThrValLeuThrProValGluVal 1904
|||||
401 GCATCCGTGTGTGCCACCGGAGAGACCGTCTGACGCCACTGGAGGGTG 450
|||||
1905 AlaIleGluAspMetGlnLysLysThrArgGluLeuAlaPheAlaThrG1 1921
|||||
451 GCCATCGAGAGACATGCAGAAAGACACGCGAGCTGGCCTTGGCACCGA 500
|||||
1921 uGlnAspProProAspAlaLysMetLeuGlnMetValLeuGlnGlySerV 1938
|||||
501 GCGAGGACCCACAGATGCTTAAGATGCTACAGATGCTGCTCAGGGCTCTG 550
|||||
1938 alG1yProThrValAsnGlnGlyProLeuGluValAlaGlnValPheLeu 1954
|||||
551 TGGGGCCCACTGAAACAGGGTCCCTGTGAGGTGGCCAGGTGTTTTTA 600
|||||
1955 AlaGluIleProGluAspProLysLeuPheArgHisAsnLysLeuArg 1971
|||||
601 GCAGAGATCCCGAAAGACCCCAAGCTCTCCGCGATCACAACAAAATTCG 650
|||||
1971 gLeuCysPheLysAspPheCysLysLysCysGluAspAlaLeuArgLys 1987
|||||
651 GCTCTGCTTCAAGACTTCTGCAGAAATGTAGGATCGCTGCGGAAC 700
|||||
1988 AsnLysAlaLeuIleG1yProAspGlnLysGluTyrHisArgGluLeuG1 2004
|||||
701 AATAAGGCCCTGATTGGGCGGACCAAGAGAGTACCAACCTGAGACTGGA 750
|||||
2004 uArgAsnTyrCysArgLeuArgGluAlaLeuGlnPro 2016
|||||
751 GCGCAACTACTTGGCCGCTGCGGAGGCTCTGCACGCC 788
|||||
seq_name: gb_est2:B1820302

```

```

REFERENCE 1 (bases 1 to 927)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11443 row: 0 column: 16
High quality sequence stop: 806.

FEATURES
     source
     1..927
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
     /clone="IMAGE:5178087"
     /clone_1lb="NIH_MGC_115"
     /lab_host="DH10B"
     /note="Organ: pooled brain, lung, testis; Vector:
     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
     source anonymous pool of 6 male brains, age range 23-27; 1
     male lung, age 27; and 1 male testis, age 69. Library is
     oligo-dT primed and directionally cloned (EcoRV site is
     destroyed upon cloning). Average insert size 1.8 kb,
     insert size range 1-3 kb. Library is normalized and
     enriched for full-length clones and was constructed by C.
     Gruber (Invitrogen). Research Genetics tracking code
     021. Note: this is a NIH-MGC Library."

BASE COUNT      214 a      257 c      292 g      164 t
ORIGIN

alignment_scores:
    quality: 1304.50      length: 329
    ratio: 4.407      gaps: 8
    Percent Similarity: 89.970      Percent Identity: 87.234

alignment block:
US-09-736-968a-2 x B1820302 ..

Align seg 1/1 to: B1820302 from: 1 to: 927

1299 G1YrLysAlaPheGluArgG1LeuSerLeuThrPheLysSerIle 1315
|||||
2 GGGAAAAGGCCCTTTGAAACCATCAACAGCCTCACATTCAAAAATTTCT 51
|||||
1315 uAspMetLysAlaArgLeuGluGluAlaIleLeuGlyThrIleG1yAlaA 1332
|||||
52 GGAATATGAAGCGCGGCTAGAGAAAGCCATTCTGGTACCATGGAGATC 101
|||||
1332 rGgGlnGluMetValArgArgSerArgGluArgSerProPheGlyAsnPro 1348
|||||
102 GACAAGAAATGGTTGGCGGAAGTCGTGAGAGAGCCCGTT. GCGAATCCG 150
|||||
1349 G1uAsnValArgTrrParGlySerValThrHisTriPlyGlnThrSerAs 1365
|||||
151 GAGAAATGTCCGCTGGCGGAAGAGCGTCCACACTGGAAAGCAAACTCAGA 200
|||||
1365 ParGValAspLysThrLysAspGluMetGluHisGluAlaLeuValGluG 1382
|||||
201 CCGGCTGGACAAAGCAAGATGAATGAAACGAGAGCCCTTGTTGGTAGG 250
|||||
1382 LysGluAlaThrGluAlaSerLeuValValLeuAspThrLeuGluIle 1398
|||||
251 GGAACCTGGCAACCGAGGCAAGCCTAGTGTCTGACACACATCGAGATC 300
|||||
1399 IleValGlnThrValMetLeuSerGluAlaArgGluSerValLeuGluL 1415
|||||
301 ATCGTCAGACGGGTGATGCTTTCAGAAACCCGGGAGAGCGCTTGGGGGC 350

```

```

seq_documentation_block: 927 bp mRNA linear EST 04-OCT-2001
LOCUS B1820302 603036864F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5178087 5',
DEFINITION mRNA sequence.
ACCESSION B1820302
VERSION B1820302.1 GI:15931852
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

1415 aValleuLysValValLeuTyrSerLeuGlySerAlaGlnSerAlaLeuP 1432
|||||
351 AGTGTGAAGCTGTGTGCTACAGCTGGGAGAGCCAGAGTCCCTCT 400
1432 heLeuGlnHisGlyLeuAlaThrGlnArgAlaLeuValSerLysPhePro 1448
|||||
401 TCTTGACACATGGCTGTGGCACCCAGAGGGCCCTGTGTCTCCAGTCCCG 450
1449 GluLeuLeuPheGluGluAspThrGluLeuGlySerAlaAspLeuLeuAr 1465
|||||
451 GAGCTGCGTGTTCGAGGAGACGAGAGCTGTGTCCGACGCTGCTCCCTAG 500
1465 GluLeuLeuArgHisCysGlySerArgIleSerThrIleArgThrHisAla 1482
|||||
501 GCTCTACGACACATGTGGCAGCCGCGATCAGACATCCGACGACGACCA 550
1482 ePAlaSerLeuTyrLeuLeuMetArgGlnAsnPheGlnIleGlyHisAsn 1498
|||||
551 GCGCTCGCTGACCTGCTCATGGCAGACAACTTCGAGATGGCCACACAC 600
1499 PheAlaArgValLysMetGlnValThr.MetSerLeuSerSerLeuVal. 1514
|||||
601 TTGCGCCCTGTGAAATGCAAGTCAACAAATGTCTCTCTCTGCTGCTGG 650
1515 GlyThrThrGlnAsnPheSerGluGluHisLeuArgArgSerLeu.LysT 1531
|||||
651 GGGACGACGACGAACTTCAGTGAAGAGCACCTGCGACCTCATCAAA 700
1531 hrIleLeuThrTyrAlaGluGluLysPheMetGlyLeuArgAspSerThrPh 1547
|||||
701 CCATCTCTACCTATGTCTGAGGAGACCACTGGGCGGAGACAGACCTT 750
1547 eAlaGluGlnValGlnAspLeuMetPheAsnLeuHisMetIleLeuThrA 1564
|||||
751 CGCAAGACGAGGTCGACGACCTGATTTCACTGCATGATTCCTGACGG 800
1564 sPThrValLysMetLys.GluHisGlnGlu.AspProGluMetLeuIleA 1580
|||||
801 ACACGCTGAAGATTCGAGGGGCCGACGAGGACCCCTGAGATT..... 843
1580 sPLeuMetLysArgIleAlaArgLysGlnGlySerProAspLeuArg 1596
|||||
844 .....GCTCTCGGCTTCCA..... 858
1597 LeuThrTrpLeuGlnAsnMetAlaGlyLysHisAlaGluLeuGlyAsnH 1613
|||||
859 .....TGTCGACATATGGCCGAGACGACGCGGAGCTGGGGAACA 899
1613 sAlaGluAlaAlaGlnCysMetVal 1621
|||||
900 CGCCGAGGCGCCGACGTGCTGTG 924
seq_name: gb_est2:BI830310
seq_documentation_block: 770 bp mRNA linear EST 04-OCT-2001
LOCUS BI830310 603073042F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5164762 5',
DEFINITION mRNA sequence.
ACCESSION BI830310
VERSION BI830310.1 GI:15941860
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 770)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.

```

```

FEATURES
    source
        cdna Library Preparation: Life Technologies, Inc.
        cdna Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
        DNA Sequencing by: Incyte Genomics, Inc.
        Clone distribution: MGC clone distribution information can be
        found through the I.M.A.G.E. Consortium/LNL at:
        http://image.llnl.gov
        Plate: LRAM11409 row: d column: 11
        High quality sequence stop: 760.
        Location/Qualifiers
            1..770
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_image="5164762"
                /clone_lib="NIH_MGC_119"
                /tissue_type="medulla"
                /lab_host="DH10B"
                /note="Organ: brain; Vector: pCMV-Sport6; Site_1: NotI;
                Site_2: EcoRV (destroyed); RNA source normal medulla from
                anonymous male age 27. Library is oligo-dT primed and
                directionally cloned (EcoRV site is destroyed upon
                cloning). Average insert size 1.3 kb, insert size range
                0.9-3 kb. Library is normalized and enriched for
                full-length clones and was constructed by C. Gruber
                (Invitrogen). Research Genetics tracking code 013. Note:
                this is a NIH_MGC Library."
BASE COUNT 164 a 239 c 223 g 144 t
ORIGIN
alignment_scores:
    Quality: 1284.00 Length: 258
    Ratio: 5.035 Gaps: 2
    Percent Similarity: 98.837 Percent Identity: 98.450
alignment_block:
    US-09-736-968a-2 x BI830310 ..
Align seg 1/1 to: BI830310 from: 1 to: 770
77 GluPheProAlaAspAspLeuGluLeuLeuGlnProArgLysCysAr 93
|||||
1 GAATGCCACGTGATGACTTGAGAGCTGCTGTCGACGCCGGGAATGCCG 50
93 gThrThrGluProGlyIleProLysAspGluLysLeuAspAlaGlnValA 110
|||||
51 GACACGACGACCCCGGAGTCCCAAGATGAGAACTGGAGACCCACAGTGA 100
110 gAlaAlaValAlaGluMetLysGlyLeuProArgGlnValPheGluGlnAspAlaSerG 126
|||||
101 GGGCCGCGGTGGAGATATATTTGAGGACTGGGTCAATTGCCACAGAGG 150
127 TyGlnTyrLeuSerAlaAlaTyrSerProValThrThrAspThrGlnAr 143
|||||
151 TATCAGTACTGAGTGCAGCATACAGCCCGCTCACACACAGACAGAGCG 200
143 gGluArgGlnLysGlyLeuProArgGlnValPheGluGlnAspAlaSerG 160
|||||
201 GGACCGACAGAAAGGGCTCCCGCCGACGCTTTGAGACAGATGCTTCTG 250
160 LysAspGluArgSerGlyProGluLysPheSerAsnAspSerArgArgGlySer 176
|||||
251 GAGACGAGAGGTCCGGCCCTGAGGACTCGAATGACTCCCGCTGGCTCG 300
177 GlySerProGluLysPThrProArgSerSerGlyAlaSerSerIlePheAs 193
|||||
301 GGCTCCCGGAGAGACACCCCTCGAAGCAGTGGTCCCTGTGACATCTCGA 350
193 PheAlaArgAsnLeuAlaAlaAspSerLeuLeuProSerLeuLeuGlnArgA 210
|||||
351 CCTGAGGAACTGCGACCTGACATCATTTGCTGCTCTGCTGAGAGCGGG 400
210 lAlaProGluLysPValAspArgArgAsnGlnThrLeuArgGlnHis 226
|||||

```

```
401 CGGCCCCAAGAGATGTGGACCGCGCAATGAACCTTGCAGCGCAGCAC 450
227 ArgProAlaLeuLeuThrLeuTyProAlaProAspGluAspGluAl 243
451 CGGCCCCCGGGCCCTGCTACCCCTCTACCCGCGCAGGATGAAC 500
243 aValGluArgCysSerArgProGluProProArgGluHisPheGlyLna 260
501 CGTGGAAACGTGTAGCCCGCTAGAGCACCCCGGAGACACTTTGGACAA 550
260 rGllLeuValLysCysLeuSerLeuLysPheGluIleGluIleGluPro 276
551 GGATCTTGCTGAAGTGTCTGTGCTCAAGTTCGAGATTGAAATGGACCCC 600
277 lIepheGlylLeuAlaLeuThrAspValArgGluLysLysLysIleSe 293
601 ATCTTTGGGACTTGGCTCTGTATGATGTGGGAGAAAAAGATGCTC 650
293 rGluAsnPheTyrrPheAspLeuAsnSerAspSer.MetLysGlyLeuLeu 309
651 GGAATAACTTCTACTTCGACCTCGAACTCGACTCCATGAAGGGGCTCTT 700
310 ArgAlaHisGlyThrHisProAlaIleSerThrLeuAlaArgSerAla.I 326
701 CGGCGTCAATGGGACCCACACCCTGCCATCTCCACCTGGCCCGCTGCCA 750
326 lPheSerValThrTyPro 332
751 TCTTCTCTGTGACTACCC 770

seq_name: gb_est2:BE531136

seq_documentation_block:
LOCUS BE531136 805 bp mRNA linear EST 09-AUG-2000
DEFINITION 601278315F1 NIH_MGC_39 Homo sapiens CDNA clone IMAGE:3610539 5',
            mRNA sequence.
ACCESSION BE531136
VERSION BE531136.1 GI:9759781
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 805)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgrabs-remail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Ling Hong/Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/MLNL at:
            http://image.jnl.gov
            Plate: LCM266 row: m column: 04
            High quality sequence stop: 770.
            Location/Qualifiers
                1..805
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:3610539"
                /clone_11b="NIH_MGC_39"
                /tissue_type="adenocarcinoma"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: pancreas; Vector: pCMV; Site_1: XhoI;
                Site_2: EcoRI; cDNA made by oligo-dT priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGACGAG(G). Library constructed
                by Ling Hong in the laboratory of Gerald M. Rubin
                (University of California, Berkeley) using ZAP-cDNA
                synthesis kit (Stratagene) and Superscript II RT (Life
```

```
BASE COUNT 169 a 246 c 249 g 141 t
ORIGIN
alignment_scores:
    Quality: 1256.00 Length: 266
    Ratio: 4.906 Gaps: 2
    Percent Similarity: 96.241 Percent Identity: 94.737
alignment_block:
US-09-736-968a-2 x BE531136 ..
Align seg 1/1 to: BE531136 from: 1 to: 805
1512 SetLeuValGlyThrThrGlnAsnPheSerGluGluHisLeuAlaArgSe 1528
|||||
2 TCCCTGTGGGGGACGACGCAAGACTTCACTGAAGACACCTCGCAGCTTC 51
1528 rLeuLysThrIleLeuThrTyAlaGluGluAspMetGlyLeuArgAspS 1545
|||||
52 ACACAAACCATCCTACCTATGCTAGAGAGACATGGGGCTGGCGGACA 101
1545 eThrPheAlaGluGlnValGlnAspLeuMetPheAsnLeuHisMetIle 1561
|||||
102 GCACCTTCGACAGACGAGCTCCAGACCTGATGTTCAACCTGCACATGATC 151
1562 LeuThrAspThrValLysMetLysGluHisGlnGluAspProGluMetLe 1578
|||||
152 CTGACGGACACCGGTGAAGATGAAGACACGAGAGACCTGAGATGCT 201
1578 uIleAspLeuMetTyrrArgIleAlaArgGlyTyrgInglySerProAspL 1595
|||||
202 CATCGACCTCATGTATACAGAAATGGCCGGGGCTACACAGGGCTACCGGACC 251
1595 euArgLeuThrTrpLeuGlnAsnMetAlaGlyLysHisAlaGluLeuGly 1611
|||||
252 TTCGGCTGACCTGGTTGCAGAACATGGCCGGGAAACACGCGAGCTGGGC 301
1612 AsnHisAlaGluAlaAlaGlnCysMetValHisAlaAlaAlaLeuValAl 1628
|||||
302 AACACGCGCGAGCGCCGCCAGTGCATTGGTGCACGGCGCCCTGTGTGCC 351
1628 agLutTyrlLeuAlaLeuLeuGluAspGlnArgHisLeuProValGlyCysV 1645
|||||
352 TGAGTACCTCGCCCTGCTGACGAGACACCGCACCTGCCCGGGCTCGC 401
1645 aISerPheGlnAsnIleSerSerAsnValLeuGluGluSerAlaIleSer 1661
|||||
402 TTTCTTCCGAACATCTCATCCAAATGTGCTTAGAGAGATCCGCCATCTCC 451
1662 AspAspIleLeuSerProAspGluGluGlyPheCysSerGlyLysHisPh 1678
|||||
452 GACGACATCTCTCGCCCGCAGACGAGAGGGCTTCTCTCGGGAGACACTT 501
1678 eThrGluLeuGlyLeuValGlyLeuLeuGluGlnAlaAlaGlyTyrrPheT 1695
|||||
502 CACTGAGCTGGGGCTGTAGGGTCTGTGAAACAGCAGCCGCGCTACTTCA 551
1695 hrMetGlyClyLeuTyrrGluAlaValaAsnGluValTyrrLysAsnLeuIle 1711
|||||
552 CCATGGGGGGGCTTACGAGGGGGTGAATGAGGCTTCAAAACCTCATTC 601
1712 ProIleLeuGluAlaHisArgAspTyrrLysLysLeuAlaAlaValHisGl 1728
|||||
602 CCCATCTCGGAAGACCCACCGTACATCAAGAAGAGTGGCGGGGTGCACGG 651
1728 yLysLeuGlnGluAlaPheThrLysIleMetHisGlnSerSerGlyTyrrG 1745
|||||
652 CAAACTCGAGGAGGGCTTACCAAGATCATGCACAGAGTTCCGGCTGGA 701
1745 lAArGValPheGlyThrTyrrPheArGValGlyPheTyrrGlyAlaHisPhe 1761
:: ::::
```



```

REFERENCE 1 (bases 1 to 777)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM1044 row: k column: 08
High quality sequence stop: 739.

FEATURES
Source
Location/Qualifiers
1..777
/organism="Mus musculus"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone_image="5005183"
/clone_lib="MCL_CGAP_Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DHI08"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 182 a 213 c 222 g 160 t
ORIGIN

alignment_scores:
Quality: 1213.00 Length: 260
Ratio: 4.720 Gaps: 2
Percent Similarity: 98.846 Percent Identity: 96.154

alignment_block:
US-09-736-968a-2 x BI078974 ..

Align seg 1/1 to: BI078974 from: 1 to: 777

1483 AAlasrleuTyrleuLeuMetArglnAsnphgIuileGlyHIsAsnph 1499
|||||
2 GCCTCCCTTACTGCTTATGCGCCAGAACTTCGAGATTGGCCATTAACCTT 51
1499 eAlaArgValLysMetGlnValThrMetSerLeuSerSerLeuValGlyT 1516
52 TCCCGGTGGAAGATGCTGTGACCATGCTCTGTCGTCCTTGCGGGA 101
1516 hTThGlnAsnPheserGluGlnHisLeuArgArgSerLeuLysThrIle 1532
|||||
102 CAACTCAGAACTTATGTAAGAGCATTTGAGAAAGTCCCTCAAGACATC 151
1533 leuThrrAlaGluGlnAspMetGlyLeuArgArgSerThrIleAlaGlu 1549
|||||
132 CTGACCTACGGAGAGGACATAGGCTGAGGAGCAGACACCTTCGCTGA 201
1549 uGlnValGlnAspLeuMetPheAsnLeuHisMetIleLeuThrAspThrV 1566
|||||
202 GCAGGTCCAGAACCTCATGATCTCAACCTGCACATGATCTGCAGACACGG 251
1566 aLysMetLysGlnHisGlnLysAspProGluMetLeuIleAspLeuMet 1582
|||||
252 TGAAGATGAGGAACACAGAGGACCTGAGATGCTATGAGACTCATG 301
1583 TyrArgIleAlaArgGlyTyrGlnGlySerProAspLeuArgLeuThrTr 1599
|||||
302 TACACAATTCGTCGGGATACCAAGGCTCTCCAGACCTGCGCGTCATG 351
1599 PLeuGlnAsnMetAlaGlyLysHisAlaGluLeuGlnLysHisAlaGluA 1616
|||||

352 GCTCAGAACATGGCTGGCAAAACATGCCAGAGCTGGCAATCATGCAAGG 401
1616 laAlaGlnCysMetValHisAlaAlaLeuValAlaGluTyrLeuAla 1632
|||||
402 CTGCCAGTGATGATGTCACAGCCGCTCCCTGGTGGCGGAATACCTTCC 451
1633 LeuLeuGlnAspGlnArgHisLeuProValGlyCysValSerPheGlnAs 1649
|||||
452 CTCCTCAGGACAGCTGCCACCTCCCTGTGGGTGTGTCTTCCGAA 501
1649 nIleSerSerAsnValLeuGlnGlnSerAlaIleSerAspIleLeuS 1666
|||||
502 TGTCTCATCCAGACGCTGCTGAGAGAGTCTGCCATCTCTGATGACATCT 551
1666 eProAspGluGlnGlyPheCysSerGlyLysHisPheThrGluLeuGly 1682
552 CGCCAGATGAGAGAGGCTTCTGCTCGGAAGAACCTTCACAGAACTGGGT 601
1683 LeuValGlyLeuLeuGlnGlnAlaAlaGlyTyrPheThrMetGlyLeu 1699
|||||
602 CTGTGTGGGCTGCTGGAGCAGGCGCGGCTACTTCACCATGGGTGCTT 651
1699 uTyrGluAlaValAsnGluValTyrLysAsnLeuIleProIleLeuGlnA 1716
|||||
652 GTATGAAGCGGTGACAGCAAGCTTACAAAAACCTTATCCCATCTTGAAG 700
1716 laHisArgAspTyrLysLysLeuAlaValHisGlyLysLeuGlnGlu 1732
|||||
701 C.CACAGAGACTACAAAGACCTGCT.GCGGTGCACGGGAACCTGCAGAG 749
1733 .AlaPheThrLysIleMetHisGln 1740
|||||
750 GGCTTCACCCAGAGATATATGACACAG 775

seq_name: gb_est2:BI181383
seq_documentation_block:
LOCUS BI181383 769 bp mRNA linear EST 10-JUL-2001
DEFINITION UNL-P-FN-ah-e-08-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
UNL-P-FN-ah-e-08-0-UNL 3', mRNA sequence.
ACCESSION BI181383
VERSION BI181383.1 GI:14655792
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 769)
AUTHORS Caetano, A.R., Johnson, R.K. and Pomp, D.
TITLE Generation and sequence characterization of a normalized cDNA
JOURNAL library from swine ovarian follicles
COMMENT Unpublished (2001)
CONTACT: Pomp, D
DEPARTMENT: Department of Animal Science
UNIVERSITY: University of Nebraska, Lincoln
ADDRESS: Lincoln, NE 68583-0908, USA
TEL: 402 472 6416
FAX: 402 472 6362
EMAIL: dpomp@unl.edu
Oligo-dt track not found, Not I site shown in beginning of sequence
is likely internal to the message.
Seq primer: M13 -29
POLYA-No.

FEATURES
Source
Location/Qualifiers
1..769
/organism="Sus scrofa"
/strain="University of Nebraska, Lincoln Swine Selection
lines"
/db_xref="taxon:9823"
/clone_image="UNL-P-FN-ah-e-08-0-UNL"
/clone_lib="UNL-P-FN"
/dev_stage="ADULT"
/lab_host="DHI08 (Life Technologies)"

```



/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site\_1: Not I; Site\_2: Eco RI; The UNL-P-FN library is a normalized library representing porcine ovarian follicles, ranging between 2.0 to 10.0 mm in diameter, collected during 7 days of the follicular phase of the pig estrous cycle. This library was derived from the library UNL-P-F2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldi, Lennon and Soares, Genome Research 6: 791-806, 1996.

TAG\_SEQ=None found"

BASE COUNT 162 a 246 c 223 g 137 t 1 others

ORIGIN

alignment\_scores:  
 Quality: 1211.00 Length: 252  
 Ratio: 4.903 Gaps: 0  
 Percent Similarity: 98.016 Percent Identity: 92.063

alignment\_block:  
 US-09-736-968a-2 x B1181383 ..

Align seg 1/1 to: B1181383 from: 1 to: 769

```

112 A1aValGluMetTyrTlEgluAspTrrValIleValHisArgTyrG1 128
|||||
11 GGGGTGGAGATGTACACCGAGCTGTGACACCGAGAGGTACCA 60
|||||
128 nTyrLeuSerAlaAlaTyrSerProValThrThrAspThrGlnArgGUA 145
1:|||||
61 GCACCTGATGCAATGTACAGCCCATCTCAGACAGACACGCGAGACC 110
|||||
145 rGgInHySG1yleuProArgGlnValPheGluGlnAspAlaSerGlyasp 161
|||||
111 GGGAGAAAGGCGCTACACCGGAGGTCTTGACAGAGATGCTTGCGGAT 160
|||||
162 GluArgSerGlyProGluAspSerAsnAspSerArgArgGlySerGlyse 178
|||||
161 GAGAGGTCTGGGCCAGAGAGACTCGGTGACCCCGCACCTCTCAGAGGCTC 210
|||||
178 rProGluAspThrProArgSerSerGlyAlaSerSerIlePheAspLeuA 195
|||||
211 CCCAGATGACACCCACGAGAGAGTGGCGCTGTGACATCTTTGACCTGA 260
|||||
195 rGAsnLeuAlaAlaAspSerLeuLeuProSerLeuLeuGlnArgAlaAla 211
|||||
261 GGAACCTGGCGGCGGACGCTCTGCTGCCCTCACTGCTGAGGCGTGGCGCC 310
|||||
212 ProGluAspValAspArgArgAsnGluThrLeuArgArgGlnHisArgPr 228
|||||
311 CCAGAGAGAGTGGAGCGGCGCAATGAGGCTCTGCGACGCGCACCGCGCC 360
|||||
228 oProAlaLeuLeuThrLeuTyrProAlaProAspGluAspGluAlaValG 245
|||||
361 CCGCGCGCTGACTACCTCTACCGACACTGATGAGAGGAGGCTGTGG 410
|||||
245 lArgGlySerArgProGluProProArgGluHisPheGlyGlnArgyle 261
|||||
411 AACGCTGACGCGCGCGGAGCCACCCGAGAAACTCTGGACACAGGATC 460
|||||
262 LeuValLysCysLeuSerLeuLysPheGluIleGluIleGluProIlePh 278
|||||
461 CTGGTCAAGTGTCTGTCCCTTAAGTTCGAGATCGAAATCGACCCATCTT 510
|||||
278 eGlyIleLeuAlaLeuTyrAspValArgGluLysLysLysIleSerGluA 295
|||||
511 TGGCATCTTAAGCCCTGTATGATGTTGGGAAAAAAGAGATCTCAGAGA 560
|||||
295 snPheTyrPheAspLeuAsnSerAspSerMetLysGlyLeuLeuArgAla 311
|||||
561 ACTTCTACTTCGACCTGACATCATGATGAGGCGTGTGCTGGCGGCC 610

```

```

312 HisGlyThrHisProAlaIleSerThrLeuAlaArgSerAlaIlePheSe 328
|||||
611 CATGACACCCACCCCTGCATTCACACCTTGCGCCGCTCTCCATCTCTCC 660
|||||
328 rValThrTyrProSerProAspIlePheLeuValIleLysLeuGluLysV 345
|||||
661 TGTACCTACCCCTCGCGCGACATCTTCTGTGTCATCAACCTGGAGAAAG 710
|||||
345 allEaGInGInGlyAspIleSerGluCysGluProTyrMetValLeu 361
711 TGTCTGACAGAGGGGNCATCAGCAGATGCTGTGACACCTACATGCTGATG 760
|||||
362 LysGlu 363
761 AAGGAG 766

```

seq\_name: gb\_est2:BG678867

seq\_documentation\_block:

LOCUS BG678867 857 bp mRNA linear EST 01-MAY-2001  
 DEFINITION 602624767F1 NCI\_CGAP\_Skn4 Homo sapiens cDNA clone IMAGE:4749689 5',  
 mRNA sequence.

ACCESSION BG678867  
 VERSION BG678867.1 GI:13910264

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

TITLE 1 (bases 1 to 857)

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: James Cleaver, M.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLNL0603 row: e column: 18

High quality sequence start: 4

High quality sequence stop: 818.

Location/Qualifiers

1..857

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4749689"

/clone\_lib="NCI\_CGAP\_Skn4"

/tissue\_type="squamous cell carcinoma"

/lab\_host="DH10B (T1 phage-resistant)"

/note="Organ: skin; Vector: pCMV-Sport6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: oligo dT. Average insert size 1.5kb. Library constructed by Life Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 209 a 254 c 247 g 147 t

ORIGIN

alignment\_scores:

Quality: 1193.00 Length: 270  
 Ratio: 4.715 Gaps: 7  
 Percent Similarity: 93.704 Percent Identity: 91.111

alignment\_block:

US-09-736-968a-2 x BG678867 ..

Align seg 1/1 to: BG678867 from: 1 to: 857

1784 HisArgLeuGluGluPheTyrThrGluArgPheGlyAspAspValValG1 1800

```

|||||
4 CACGGCGTGGAGAGTTCTACACGAGAGATTGGCGACGCGTGTGA 53
1800 uillellysaspserlyrprovalasplyserslyleuaspsenl 1817
54 GATTATCAAAAGACTTAACCTGTGGCAAGTCCAACTTGACTGACACAAA 103
1817 ysalaTyrlleGlInleThrTyrlValGlupProTyrlPheaspThrTyrlGlu 1833
104 AGGCTTACATCCAGATCACGATATGGAACCGTACTTGTACTTACGAG 153
1834 leulysaspargValThrTyrlPheaspArgasnTyrlGleuArGThrph 1850
154 CTCAAGAGACGGGTGACCTACTTGGACCGCAACTATGGCTTGCACATT 203
1850 eleuPheCysThrProPheThrProaspGlyArgAlaHisGlyGluLeuP 1867
204 CTTCTTTCGACGCGCTTCACGCGGATGGCGCGACAGGGGAGCTGC 253
1867 roGlInHisLysArgLysThrLeuLeuSerThrasPHisAlaPhePro 1883
254 CCGAGCAACAACAGCTAAGACGCTGCTCAGCAACGACGCGCTTCCC 303
1884 TyrlleLysThrArglleArgValCysHisArgGluGluThrValLeuTh 1900
304 TACATCAAGACTCGCATCCGTTGTGCCACCGGAGAGAGCGGTGTGAC 353
1900 rProValGluValAlaIleGluAspMetGlnLysLysThrArgGluLeuA 1917
354 GCCAGTGGAGGTGGCCATCAAGACATGCAAGAAAGAACAGGGAAGCTGG 403
1917 laPhealarThGluGluAspProAspAlaLysMetLeuGlnMetVal 1933
404 CCTTGGCACCGACGAGACCCACAGACTCTAAGATGCTACAGATGTG 453
1934 leuGlnLysSerValGlyProThValAsnGlnGlyProLeuGluValAl 1950
454 CTTAGGGGCTGTAGGGCCACCGTGAACCA.GGTCCCTGGAGGTGGC 502
1950 agInValPheLeuAlaGluIleProGluAspProLysLeuPheArgHis 1967
503 CCAGCTGTTTTAGCAGAGATCCGGGAAGACCCCAAGCTTCCGGATC 552
1967 lAsanLysLeuArgLeuCys.PheLysAspPheCysLysLysCysGluAs 1983
553 ACAACAAATTGGCGCTCTCTTCAGAGACTTCTGCAGAAATGTGAGGA 602
1983 pAlaLeuArgLysAsnLysAlaLeuIleGlyProAspGlnLysGluTyH 2000
603 TGGCTTCGGGAAAAATTAAGCCCTGATGGGGCGGACCAAGAGATACC 652
2000 lAsArg.GluLeuGlu.ArgAsnTyrlCysArgLeuArg.GluAlaLeuGln 2015
653 ACCGGGAGGCTTGAAAGCGCAACTACGCCCTGGGGAGGCTTGAGAG 702
2016 ProLeuLeuThrgIn...ArgLeuProGlnLeuMetalaProThrPro.. 2030
703 CCCGCTTACCCCAAGCGGCGCTGCCCACTAATGGCAACCCACCCCA 752
2031 .ProGlyLeuArgAsnSerLeu.AsnaArgAlaSerPheArgLysAlaAsp 2046
753 GCCAGGCTTTCAGAGACTCTTGGACACAGAGCAAGTTTCCCGAAGGGAG 802
seq_name: gb_est2:BI159195
seq_documentation_block:
LOCUS BI159195 778 bp mRNA linear EST 05-JUN-2001
DEFINITION 602922704P1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5063064 5',
ACCESSION BI159195
VERSION BI159195.1 GI:14619196
KEYWORDS EST.
SOURCE house mouse.

```

```

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 778)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lotmar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M1170 row: g column: 01
High quality sequence stop: 750.
location/Qualifiers
1..778
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5063064"
/clone_lib="NIH_CGAP_Mam3"
/tissue_type="tumour, gross tissue"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1; NotI;
Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert 2 kb. Library constructed by Life
Technologies, catalog #12017-018. Investigators providing
samples: Lotmar Hennighausen/Chu-Xia Deng, NIH Reference
for transgenic model: Xu et al., Nature Genetics 22, 37-43
(1999). Note: this is a NIH_CGAP Library."
BASE COUNT 173 a 223 c 226 g 156 t
ORIGIN
alignment_scores:
Quality: 1178.00 Length: 262
Ratio: 4.656 Gaps: 4
Percent Similarity: 96.565 Percent Identity: 93.511
alignment_block:
US-09-736-968a-2 x BI159195 ..
Align seg 1/1 to: BI159195 from: 1 to: 778
1373 GUMetGluHisGluAlaLeuValGluLysAsnLeuAlaThrgAlaSe 1389
|||||
1 GAAATGAGACATGAAAGCTGTGTGAGCGGAACCTGGCGACAGAGCCAG 50
1389 rLeuValAlaLeuAspThrLeu.GluIleIleValGlnThrValMetLeu 1405
51 CCTGTGTGTGTGAGACAGCTGTGAGACATCGTCGCAACAGTGTGTG 100
1406 SerGluAlaArgGluSerValLeuGluValAlaLeuLys.ValValLeuT 1422
101 TCCGAGGCCCGCGAGAGACTCTGTAGTCCCGTGCAGAAATGTTGCTCCT 150
1422 ySerLeuGlySerAlaGlnSerAlaLeuPheLeuGln.HisGlyLeuAl 1438
151 ACAGTCTTGGAGAGCCCGAGAGTGCCTGTCTCTGCAACATGGGCTGGC 200
1438 aThrgInArgAlaLeuValSerLysPheProGluLeuLeuPheGluGlu 1455
201 CACCGAGCGGGCGCTGCTCCAAAGTTCCGAGAGCTGTTTCAGAGAG 250
1455 sPThrGluLeuCysAlaAspLeuCysLeuArgLeuAlaHisCysGly 1471
251 ACACGAGACTTGTGTCGACCTGTGCTGAGAGACTTCTGCAACTGTGGC 300
1472 SerArgIleSerThrIleArgThrHisAlaSerAlaSerLeuTyrlLeu 1488
|||||

```

```

301 AGCCGCATCAGCAKCCATCCGCATGCAGCCGACGGCTCCCTCACTACGTCT 350
1468 UMEtArgGlnAspNheGulileGlyHisAspNheAlaArgValIlysmetG 1505
351 TATGGGCGAGAACTTCAGATTGGCCATACCTTCCCTCGTGTGAAGATGC 400
1505 InValThrMetSerLeuSerSerLeuValGlyThrThrGlnAspPheSer 1521
401 TGTGTACCATGTCTCTGTGCTCCCTTGTGTGGGACAACCTCAGACTTTAGT 450
1522 GUGUGUHTSteuATgArgSerLeuIlysthrIleLeuThrTyRAlaGluG 1538
451 GAAGAGCATTTGAGAAAGTCCCTCAAGACCACTCAGACTCGCAGAGAGA 500
1538 uAspMetIyLeuArgAspSerThrPheAlaGluGlnValGlnAspLeuM 1555
501 GGACATAGGGGTGAGGAGACAGACTTCCTGCTGAGCGAGGTCCAGGACTCA 550
1555 eLpHeAspLeuHisMetIleLeuThrAspThrValIlysmetGlyGluHis 1571
551 TGTTCACCTGTCACATGATCTTGACAGACAGCGTGAAGAATGAGGAACAC 600
1572 GlnGluAspProGluMetLeuIleAspLeuMetTyRArgIleAlaArgG 1588
601 CAGGAGGACCTTGATGATGTCATGAGACCTCATGATGATGATGATGCTGG 650
1588 yTyRGlnGlySerProAspLeuArgLeuThrTrpLeuGlnAspMetAlaG 1605
651 ATACCAGTCTCTCCAGACTGCGGCTGACCTGCGTGGCAGAACATGAGCTG 700
1605 IyIyHisAlaGluLeuGlyAsnHisAla.GluAlaAlaGlnCysMetVa 1621
701 GCAAAACATGCGAGAGCTGGGCATCATGCAAGAGCGTCCCATGATGATGT 750
1621 HisAlaAlaAlaLeuValAlaGluTyR 1630
751 CCACGCGCTGCTCCCTGGGTGGCGAATAC 778

seq_name: gb_est1:AW742035

seq_documentation_block:
LOCUS AW742035 692 bp mRNA linear EST 07-SEP-2000
DEFINITION up52c04.y1 Soares mouse NMIE Mus musculus cDNA clone IMAGE:2779686
5' similar to WP.F46H5.4 CE04590 RAT TRG GENE PRODUCT ;, mRNA
sequence.
ACCESSION AM742035
VERSION AM742035.1 GI:7653891
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 692)
JOURNAL NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Other_ESTs: up52c04.x1
Contact: Robert Strausberg, Ph.D.
Email: cgrabs-r@mail.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: 40RP from GlpDco
High quality sequence stop: 473.
Location/Qualifiers
1..692
FEATURES
SOURCE
/organism="Mus musculus"
/strain="C3H x 101 (F1 stock)"
/db_xref="taxon:10090"
/clone="IMAGE:2779686"
/clone_lib="Soares_mouse_NMIE"
/sex="male"
/dev_stage="newborn"

```

```

/lab: host="DRI0B"
/Note="Organ: inner ear, 170 pooled; Vector: pT73D-Pac;
Site_1: Not 1; Site_2: Eco RI; 1st strand cDNA was primed
with a Not I- oligo(dT) primer [5'-
TGTATCCATCTGTAAGTGGAGGCGCGCCGACACCTATTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
is normalized, and was constructed and donated by Bento
Soures and M. Fatima Bernaldo (University of Iowa) and R.
Hardisty, A. Varela-Carver, P. Mburu and S.D.M. Brown (MRC
UK Mouse Genome Centre and Mammalian Genetics Unit,
Harwell, UK)."
BASE COUNT      183 a      186 c      188 g      133 t      2 others
ORIGIN
alignment_scores:
    Quality: 1151.00      Length: 230
    Ratio: 5.161      Gaps: 0
Percent Similarity: 96.957      Percent Identity: 94.348
alignment_block:
US-09-736-968A-2 x AW742035      ..
Align seg 1/1 to: AW742035 from: 1 to: 692
1778 TysLeuAlaGluIleSerHisArgLeuGluGluPheTyrThrGluAlaGph 1794
3 AAGCTTGCAGAGATCTCACACGGCTGGAGGAGTCTTATACGGAAAGGTT 52
1794 eGlyAspArgValAlaGluIleLeuLysAspSerTyrProValAspLysS 1811
53 CGGGAGATGATGTGGTAGAGATCATCAAGATTCCTAACCGAGTGACAGT 102
1811 erLysLeuAspSerGlnLysAlaTyrIleGlnIleThrTyrValGluPro 1827
103 CCAGACTGGACCCACAGAGAGCGTACATACAGATAACCTATGTGGAGCA 152
1828 TyrPheAspThrTyrGluLeuLysAspArgValThrTyrPheAspArgAs 1844
153 CATTTCCGACACTATGAGCTCAAGAGATCGGAGTACCTACTGTGATCGGAA 202
1844 nTyrGlyLeuArgThrPheLeuPheCysThrProPheThrProAspGlyA 1861
203 CTATGGGTGGGGGCTTCCTCTTGTGCACACCTTCACACCGATGGAGC 252
1861 rGAlaHisGlyLeuProGluGlnHisLysArgLysThrLeuLeuSer 1877
253 GTGCCACGAGAGAGTTGGCCGAACACACAAACGCAAGCGCTGTGACG 302
1878 ThrAspHisAlaPheProTyrIleLysThrArgIleArgValCysHisAr 1894
303 ACGGAGCATGCGCTTTCCTACATCAAGACACAGCATCCAGTGTGCCACCG 352
1894 gGluGluThrValLeuThrProValGluValAlaIleGluAspMetGlnL 1911
353 TGAGGAGACAGTGGCTGACACCACTGGAGGAGGCGCATTTGAGGACATG 402
1911 yLysThrArgGluLeuAlaPheAlaThrGluGlnAspProProAspAla 1927
403 AGAAGACCCGGGAGCTGGCTTTGCCACCGACGAGAGACCTCTCAATGCC 452
1928 LysMetLeuGlnMetValLeuGlnGlySerValGlyProThrValAsnGl 1944
453 AAGATGCTGCAGATGTTCTCCAGGGTCTGTGGGACCCACACTGTGAACCA 502
1944 nGlyProLeuGluValAlaGlnValPheLeuAlaGluIleProGluAspP 1961
503 NGTGTCCCTTGGAAATGGCCANGTGTTTGTTCAGAGATCCAGAAAGATC 552
1961 rOLysLeuPheArgHisHisAsnLysLeuArgLeuCysPheLysAspPhe

```

553 CCAAGCTCTTCGACATCACAACTCCGGCTCTGCTTCAAGATTTC 602  
1978 CysIysIysCysGluAspAlaLeuArgIysAsnIysAlaLeuIleGIYPr 1994  
|||||  
603 TGCATAAAGTGCAGAGATGCACTGAGAAAGAACAAAGCCCTGATTGGGCC 652  
1994 oAspGlnIysGluTyrHisArgGluLeuGluArgAsnTyr 2007  
|::|  
653 CAACCCAGATGATACCAACCGGAGAGCTTGAGCTCACTAT 692



CC useful for detecting the CLASP-7 polypeptide. (ii) is useful for  
CC inhibiting (i) by recombinant methods. (i) or its fragment are useful for  
CC pharmaceutical composition (C), comprising (i) or (ii), can be useful  
CC for treating CLASP-7-mediated disease such as an autoimmune disease  
CC caused or exacerbated by increased activity of T helper cells. Autoimmune  
CC diseases which can be treated using (C) include multiple sclerosis,  
CC juvenile diabetes and rheumatoid arthritis. (i) is useful for treating  
CC toxemia or pregnancy induced hypertension, pruritic urticarial papules  
CC and Rh incompatibility. (i) is also useful as a diagnostic reagent for  
CC immune and other disorders, since diseases characterised by  
CC overproduction or depletion of lymphocytes in blood or other organs may  
CC be detected by monitoring the level of (i) or its mRNA. CLASP-7 has been  
CC mapped to the chromosomal location 19q13.2. The present sequence  
CC encodes the human CLASP-7 protein as given in the present invention.  
XX  
SQ

Sequence 6372 BP: 1273 A; 2050 C; 1849 G; 1200 T; 0 other;

alignment\_scores:

Quality: 10569.00	Length: 2047
Ratio: 5.163	Gaps: 0
Percent Similarity: 100.000	Percent Identity: 100.000

alignment\_block:

US-09-736-968a-2 x AAH43851 ..

Align seg 1/1 to: AAH43851 from: 1 to: 6372

```
1  MetAlaLAserGIuArGaRgaLaPheAlaHIsLysIleAsnArGThrVa 17
   |||||||
13  ATGGCTGCTCCGAGCGCGCGCTTCGCGCACAAAGATCAACAGACGCGT 62
   |||||||
17  LAIAAlaGIuValArGLysGlnValSerArGluArGlySerProH 34
   |||||||
63  GGCGGAGAGGTGCGGAGAGGTGTCCGGGAACGACGTGCTCCGCC 112
   |||||||
34  IsSerArGaRgCySerSerSerLeuGIuValProLeuThrGluVal 50
   |||||||
113  ACTCCAGACAGCGCTGCAGACGCTCCCTGGGGTCCCACTGACTGAAGT 162
   |||||||
51  ValGIuProLeuAspPheGIuAspValLeuLeuSerArGProProAspAl 67
   |||||||
163  GTCGAGGCCCTTGACATTGAGATGTACTTCAGCGCGGCCACCAAGATGC 212
   |||||||
67  aGIuProGIyProLeuArGAspLeuValGIuPheProAlaAspAspLeuG 84
   |||||||
213  TGAAGCCGGGGCCCTCAGGACCTGGTAAGATTCCAGCTGATGACTTGG 262
   |||||||
84  IuLeuLeuGIuGIuProArGIuCyArGThrThrGIuProGIyLeuPro 100
   |||||||
263  AGCTGCTGCTGCAGCCCGGGGATGCCGGACACAGGAGCCGGGATCCCC 312
   |||||||
101  LysAspGIuLysLeuAspAlaGlnValArGAlaAlaValGIuMetTyrI1 117
   |||||||
313  AAGGATGAAAAAAGCTGATGCCAGTGAGGGCCGCGTGAGATGTATAT 362
   |||||||
117  eGIuAspTrValIleValHIsArGArGTyrGIuTyrLeuSerAlaAlaT 134
   |||||||
363  TGAAGGATGGGTATTGTCACAGAAAGTATCACTGAGTGCAGCAT 412
   |||||||
134  YrSerProValIThrThrAspThrGlnArGGIuArGGIuLysGIuLeuPro 150
   |||||||
413  ACAGCCCCGTCACACAGACACACAGCGGAGGAGACAGAAAGGCTCCCC 462
   |||||||
151  ArgGIuValPheGIuGlnAspAlaSerGIyAspGIuArGSerGIyProGI 167
   |||||||
463  CGCCAGGTCTTTGAGCAGATGCTTCTGAGACAGAGAGTCCGGCCCTGA 512
   |||||||
167  uAspSerAsnAspSerArGArGIySerGIySerProGIuAspThrProAl 184
   |||||||
513  GGACTGCAATGACTCCCGGCGTGGCTCGGGCTCCCGGAAAGACACCCCTC 562
```

```
184  rGSerSerGIyAlaSerSerIlePheAspLeuArGAsnLeuAlaAlaAsp 200
   |||||||
563  GAACAGAGTGGTCCCTAGCATCTTCGACCTGAGAGAACCTGGCAGCTGAC 612
   |||||||
201  SerLeuLeuProSerLeuLeuGIuArGAlaAlaProGIuAspValAspAr 217
   |||||||
613  TCATTGCTGCCCTCTTCGCTAAGACGGGGGGCCCCCAAGATGTGGACGG 662
   |||||||
217  GArGAsnGIuThrLeuArGArGIuHIsArGProProAlaLeuLeuThrL 234
   |||||||
663  GCGCAATGAAGAACCTTGACGCGCAGCACCGGGCCCCCTGCTGACACC 712
   |||||||
234  eUTyrProAlaProAspGIuAspGIuAlaValGIuArGySerArGPro 250
   |||||||
713  TCTACCCGGGACCTGACGAGATGAACCGGTGAACCTGTAGCGGCCCA 762
   |||||||
251  GIuProProArGIuGIuHIsPheGIuGIuArGIleuValLysCysLeuSe 267
   |||||||
763  GAGCCACCCCGCGAGCACTTGGACAAAGATCTTGGTCAAGTGTCTGTC 812
   |||||||
267  rLeuLysPheGIuIleGIuIleGIuProIlePheGIyIleLeuAlaLeuT 284
   |||||||
813  GCTCAAGTTCGAAGATTGAATGAGCCCATCTTGGGATCTGGCTCTGT 862
   |||||||
284  YrAspValArGGIuLysLysLysIleSerGIuAsnPheTyrPheAspLeu 300
   |||||||
863  ATGATGTGCGGGAGAAAAAGAAATCTCGAGAACTTCACTTCGACCTG 912
   |||||||
301  AsnSerAspSerMetLysGIyLeuLeuArGAlaHIsGIyThrHIsProAl 317
   |||||||
913  AACTCGGACTCCATGAAGGGGCTCTTCGGGCTCATGACACCCACCTGCG 962
   |||||||
317  aIleSerThrLeuAlaArGSerAlaIlePheSerValThrTyrProSerP 334
   |||||||
963  CATCTCCACCTTGCGCCGCTGCGCATCTTCTGTGACCTACCCCTCAC 1012
   |||||||
334  rOAspIlePheLeuValIleLysLeuGIuLysValLeuGlnGlnGIyAsp 350
   |||||||
1013  CTGACATCTTCCTGCTCATCAAGTTGGAGAGGTGCTTCAGCAAGGGGAG 1062
   |||||||
351  IleSerGIuCyGyGIuProTyrMetValLeuLysGIuValAspThrAl 367
   |||||||
1063  ATCACTGATGCTGTGACCTTACATGGGTGTGAAGAGTGGACACACACC 1112
   |||||||
367  aLysAsnLysGIuLysLeuGIuLysLeuArGLeuAlaAlaGIuInPheC 384
   |||||||
1113  CAAGAACAAAGAGAAAGCTAGAGAAAGCTCGCGCGCGCAGAGTTCT 1162
   |||||||
384  ySThrArGLeuGIyArGTyrArGmetProPheAlaIATrPThAlaValHIs 400
   |||||||
1163  GCACCCGCGCTGGCGCGCTACCGCATGCCCTTCGCTGACCGCGCTGCAC 1212
   |||||||
401  LeuAlaAsnIleValSerSerAlaGIyGIuLeuAspArGAspSerAspSe 417
   |||||||
1213  TTGGCAACATGTGAGCAGCGCTGGGAGCTGACCGGGAGCTGTGACTC 1262
   |||||||
417  rGIuGIyGIuArGArProAlaIATrPThAspArGArGArGIyProG 434
   |||||||
1263  GGAAGGCGAGCGCGCGCGCAGCTGACAGACCGCGCGCTGGGGGCCCC 1312
   |||||||
434  InAspArGAlaSerSerGIyAspAspAlaCySerPheSerGIyPheArG 450
   |||||||
1313  AGGACCGGGCGAGTArGTGGGAGCAGCGCTTCACCTTCCTGCGTTCCGT 1362
   |||||||
451  ProAlaThrLeuThrValIThrAsnPhePheLysGlnGIuAlaGIuArGLe 467
   |||||||
1363  CGAGCAGCGCTACTGTCAAAACTTCTTTAAGCAGAGAGGTGAGCGACT 1412
   |||||||
467  uSerAspGIuAspLeuPheLysPheLeuAlaAspMetArGArGProSerS 484
   |||||||
1413  CAGTGACGAGGACGCTCTCAAGTTCTCTGCTGACATGAGCGCGCGCTGT 1462
   |||||||
484  eTLeuLeuArGArGLeuArGProValIThrAlaGlnLeuLysIleAspIle 500
```

|||||  
1463 CCCGTCGGCGGCACTACGTCTGTGACTGCCCCAGCTCAGATGCAACTT 1512  
501 SerProAlaProGluAsnProHisPheCysLeuSerProGluLeuLeuH 517  
|||||  
1513 TCCTCGGCTCTCGAAAAATCCCACTTCTGCTCTCCCGCTGAGTGGCTCA 1562  
517 sIleLeuProTyrProAspProArgGlyArgProThrLysGluIleLeuG 534  
|||||  
1563 TATCAACGCCCTACCCGGACCCAGGGGGCCGCCACCACAAAGAGATTCTGG 1612  
534 IupheProAlaArgGluValTyrAlaProHisThrSerTyrArgAsnLeu 550  
|||||  
1613 AGTTCGCCCGCCCGGAAGTCTATGCCCCCATATCCAGCTACAGGAACCTG 1662  
551 LeuTyrValTyrProHisSerLeuAsnPheSerSerArgGlnLysEryA 567  
|||||  
1663 CTGTACGTGTACCCGCGACAGCTCAACTTCAGCAGCGCGCAGGGCTCCGT 1712  
567 IArgAsnLeuAlaValArgValGlnTyrMetThrGlyLysAspProSerG 584  
1713 GCGCAACCTTGCTGTGTGGAGTGCATGACATGACAGCGAGAGCCCAAGCC 1762  
584 IuAlaLeuProValIlePheGlyLysSerSerCysSerGluPheThrArg 600  
|||||  
1763 AGGCTCGCGGGTCATCTTTGGCAAGTCCAGCTGCAGTGAATTTACCCGC 1812  
601 GluAlaPheThrProValValTyrHisAsnLysSerProGluPheTyrG 617  
1813 GAGGCTTCACACCGGTGGTGTCTACCATAAACAAGTCCCGCGATTACGA 1862  
617 uGluPheLysLeuHisLeuProAlaCysValThrGluAsnHisLeuL 634  
1863 GGAAGTTCAACTGATCTTCCAGCTCCGCGACAGAAACATCACCTGCG 1912  
634 eupheThrPheTyrHisValSerCysGlnProArgProGlyThrAlaLeu 650  
|||||  
1913 TGTTCACCTTCTACCAATGTCAGCTGCCAGCCCGCGGGACACTGCCCTG 1962  
651 GluThrProValGlyPheThrTrpIleProLeuLeuGlnHisGlyArgLe 667  
1963 GAGACACCCCGTGGGCTTACTTGATGCCACTGCTGCAGACGCGGCGCT 2012  
667 uArgThrGlyProPheCysLeuProValSerValAspGlnProProP 684  
2013 GAGGACCGGGCCCTTCTGTCTCCAGATGTCGTGGACCAAGCCGCCCA 2062  
684 eTyrSerValLeuThrProAspValAlaLeuProGlyMetArgTrpVal 700  
2063 GCTATTCGGTGTCTACACCGCATGTGGCGCTTCGGGCGATGCCCTGGGTG 2112  
701 AspGlyHisLysGlyValPheSerValGluLeuThrAlaValSerEryA 717  
2113 GACGGTCACAAAGGGCGTGTTCAGTGTGAGCTCAGACAGCGGTGCTCTGT 2162  
717 IHisProGlnAspProTyrLeuAspLysPhePheThrLeuValHisValL 734  
|||||  
2163 GCACCCCCAGACCCCTACCTGGACAATTCCTCACCCGTGTGACAGTCTCC 2212  
734 euGluGluGlyAlaPheProPheArgLeuLysAspThrValLeuSerGlu 750  
2213 TGGAGGAGGAGGAGCTTCCCATTCGGGCTCAAGGACACTGTGCTGAGCGAG 2262  
751 GlyAsnValGluGlnLeuLeuArgAlaSerLeuAlaAlaLeuArgLeuAl 767  
2263 GGCACAGTGGAGCA3GAGCTGGCGGCGCAGTCTTGCAAGCACTGGCGCTGC 2312  
767 aserProGluProLeuValAlaPheSerHisHisValLeuAspLysLeuV 784  
2313 CAGCCCGCGAAACCCCTTGTCCTTCGCCACCAAGCGTGGAGCAAGCTCG 2362  
784 aIArgLeuValIleArgProIleIleSerGlyGlnIleValAsnLeu 800  
|||||

2363 TCCGTCGTGTCATCAAGCCCCCGATCATCAGTGGCCAGATTGTGAACTGTG 2412  
801 GlyArgGlyAlaPheGluAlaMetAlaHisValValSerLeuValHisAr 817  
2413 GGCCTGGAGAGCTTTGAAGCAATGGCCCATGTAGTACACCTTGTTCACCG 2462  
817 gSerLeuGluAlaAlaGlnAspAlaArgGlyHisCysProGlnLeuAlaA 834  
2463 GAGCCTGGAGGAGCCCAAGATGCCCGGGGTCACTGCCACAGCTGGCTGT 2512  
834 IaTyrValHisTyrAlaPheArgLeuProGlyThrGluProSerLeuPro 850  
2513 CCTACGTCACATACCTCTTGGCTTCTCCGTGGCACTGAGCCAGCCCTCCCG 2562  
851 AspGlyAlaProProValThrValGlnAlaAlaThrLeuAlaArgLysE 867  
2563 GATGGAGGCCCTTCAGTGTACAGTGCAGGCTGCACACTGAGCCCGGTGCTT 2612  
867 rGlyArgProAlaSerLeuTyrLeuAlaArgSerLysSerLleSerSerS 884  
2613 TGGTGGCCCCGCAACCTCTACCTGGCGCGTTCCAAGAGCATCAGCACCA 2662  
884 eArgAsnProAspLeuAlaValAlaProGlySerValAspAspGluValSer 900  
2663 GCAACCCGTACCTCCCGCTGGCCCTGCGCTCTGTGATGACAGAGTTTCC 2712  
901 ArgIleLeuAlaSerLysLeuLeuHisGluGluLeuAlaLeuGlnTrpVa 917  
2713 CCATTCCTGGCCAGCAAGCTGCTTCACGAGAGAGTGGCTGTGCAAGTGGGT 2762  
917 IValSerSerSerAlaValArgGluAlaIleLeuGlnHisAlaTrpPheP 934  
2763 GGTCAAGACAGTACGCGGTACGCGAGGCCATCTCCAGACACCCGTGGTCT 2812  
934 hePheGlnLeuMetValLysSerMetAlaLeuHisLeuLeuGlyGln 950  
2813 TTTTCCAGCTCATGTGTGAAGATATGGCGCTGCACCTCTGCTGTGGCCAG 2862  
951 ArgLeuAspThrProArgLysLeuArgPheProGlyArgPheLeuAspAs 967  
2863 GCAGTAGACACACCCCGCAAGCTGGCTTCCCGGAGACCTTCCTGAGAGA 2912  
967 rIleThrAlaLeuValGlySerValGlyLeuGluValIleThrArgValH 984  
2913 CATCACTGCCTTGGTGGGCTGTGTGGGCTGAGGTCAATCACCCGTGTGCC 2962  
984 IsLysAspValGluLeuAlaGlnHisLeuAsnAlaSerLeuAlaPhePhe 1000  
2963 ACAAGAGATGTGAGCTGGCCGAGCACCTCAAGCCAGCCCTGGCTTTCTTC 3012  
1001 LeuSerAspLeuLeuSerLeuValAspArgGlyPheValPheSerLeuVa 1017  
3013 CTCAGTGAACCTTCTGTCCCTGTGGAGACCGGGGCTTGTCTTCACCGTGT 3062  
1017 IArgAlaHisTyrLysGlnValAlaThrArgLeuGlnSerSerProAsn 1034  
3063 CCGGGCCCACTACAAGCAGGTGGCCACGCGGCTCCAGTCCGCTCCATATC 3112  
1034 roAlaAlaLeuLeuThrLeuArgMetGluPheThrArgIleLeuCysSer 1050  
3113 CAGCAGCCCTGTGACCTTGGCATGTGAATTCACCCGATCTGTGACAGC 3162  
1051 HisGlnHisTyrValThrLeuAsnLeuProCysCysProLeuSerProp 1067  
3163 CACGAGCACTAGTGAACCTCAACCTCCCGTGGCCCGCTGTACCTCC 3212  
1067 oAlaSerProSerProSerValSerSerThrThrSerGlnSerSerThrp 1084  
3213 AGCCTCGGCTTCCCTCTGTGTCTCCACACCACTCCACAGTGCACACT 3262  
1084 heserSerGlnAlaProAspProLysValThrSerMetPheGluLeuSer 1100  
3263 TCTCCAGCCAAAGCCCGGAGCCCAAGGTGACCAAGCATGTTCGAATGAGT 3312

1101 GlyProPheArgGlnGlnHisPheLeuAlaGlyLeuLeuThrGluLe 1117  
|||||  
3313 GGACCATTCGGCAGACAGCACTCTAGCTGGCTCCTGCGAGGAGCT 3362  
1117 uAlaLeuAlaLeuGluProGluAlaGluGlyAlaPheLeuLeuHisLysL 1134  
|||||  
3363 GGCACTGGCCCTCGAACCTGAGGCTGAAGGGGCAATCCTGTTGCAAGA 3412  
1134 ysaAlaIleSerAlaValHisSerLeuLeuGlyHisAspThrAspPro 1150  
|||||  
3413 AGGCCATCAGTGTGTGCACAGCCTGCTATGTGSCCATGACACGACCCC 3462  
1151 ArgTyrAlaGluAlaThrValLysAlaArgValAlaGluLeuTyrLeuPr 1167  
|||||  
3463 CGCTACGCCGAGGCCACTGTGAAAGGCTGTGTCGCCGAGCTGTACCTGCC 3512  
1167 oLeuLeuSerIleAlaArgAspThrLeuProArgLeuHisAspPheAlaG 1184  
|||||  
3513 ACTGCTTCGATGTGCAGGATACCTTGGCCAGGCTGCATGACTTTCTGTG 3562  
1184 IugIyProGlyIleArgSerArgLeuAlaSerMetLeuAspSerAspThr 1200  
|||||  
3563 AGGGCCCAAGTGCAGGGGTCAAGACTGGCTCAATGCTTGACTGCAGACAA 3612  
1201 GluGlyIugIyAspIleAlaGlyThrIleAsnProSerValAlaMetAl 1217  
|||||  
3613 GAAGCCGAGGGGACATTCGGGTACCATCAACCCCTGTGGGCCATGCGC 3662  
1217 aIleAlaGlyGlyProLeuAlaProGlySerArgAlaSerIleSerGln 1234  
|||||  
3663 CATTCGTGTGGCCCTCAGCCCTGTGGCTCCGGGCCAGCATCTCCCAAG 3712  
1234 IyProProThrAlaSerArgAlaGlyCysAlaLeuSerAlaGluSerSer 1250  
|||||  
3713 GGCCACCAACGGCTTCTGCGCAGCGTGTGCCCTCTGCTGAGTCAAGC 3762  
1251 ArgThrLeuLeuAlaCysValLeuTyrValLeuLysAsnThrGluProAl 1267  
|||||  
3763 CGGACCTTGTGGCGTGTGTGCTGTGGTGTGTAACCAACCGGCGGCGC 3812  
1267 aLeuLeuGlnArgTyrPalaThrAspLeuThrLeuProGluLeuGlyArgL 1284  
|||||  
3813 GCTCTCGAGGGCTGGGCGCACTGACCTGACACTCCCCAGCTGGGAGCTC 3862  
1284 euleuAspLeuLeuTyrLeuCysLeuAlaAlaPheGluTyrLysGlyLys 1300  
|||||  
3863 TGTGGACTTGTCTGTACTTGTCTGTGCTGTGCTTGAAGTACAAAGGGAAA 3912  
1301 LysAlaPheGluArgIleAsnSerLeuThrPheLysLysSerLeuAspMe 1317  
|||||  
3913 AAGGCTTGTGAACGATCAACAGCTCACAATCAAAAATCTGTGGATAT 3962  
1317 tLysAlaArgLeuGluGluAlaIleLeuGlyThrIleGlyAlaArgGlnG 1334  
|||||  
3963 GAAGGCGGGGTAGAGGAAGCATTTGGGTACATCGAGACTGAGAACG 4012  
1334 IuMetValArgArgSerArgGluArgSerProPheGlyAsnProGluAsn 1350  
|||||  
4013 AAATGTGTTCCGGAGAGTGTGAGAGAGCCGTTTGGGAATCCAGAGAAC 4062  
1351 ValArgTyrPArgLysSerValThrHisTyrPysGlnThrSerAspArgVa 1367  
|||||  
4063 GTGGCGCTGGCGAAGAGGCTACACACTGGGAAGCAAACTTAGACCGCGT 4112  
1367 IAspLysThrLysAspGluMetGluHisGluAlaLeuValGluGlyAsnL 1384  
|||||  
4113 GGACAAGACCAAGATGAATGAACAAGAGCCTTGTGTGAAGAGAAC 4162  
1384 euAlaThrGluAlaSerLeuValValLeuAspThrLeuGluIleIleVal 1400  
|||||  
4163 TGGCAACGAGGCAAGCCTAGTCTGTCTGACACACTGAGATCATCTGTG 4212

1401 GlnThrValMetLeuSerGluAlaArgGluSerValLeuGlyAlaValLe 1417  
|||||  
4213 CAGACGGTATGCTTTCAGAAAGCCCGGAGAGCGTCTTGGGGGAGTGT 4262  
1417 uLysValValLeuTyrSerLeuGlySerIleGlnSerAlaIlePheLeuG 1434  
|||||  
4263 GAAGTGTGTGCTGTACAGCTGGGCGAGTGCACAGAGTGCCTTCTTTC 4312  
1434 IuHisGlyLeuAlaThrGlnArgAlaIleValSerLysPheProGluLeu 1450  
|||||  
4313 ASCATGGCTGGGCGACCCAGAGGGCCCTGTGTGCCAAGTTCGGGAGCTG 4362  
1451 LeuPheGluGluAspThrGluLeuCysAlaAspLeuCysLeuArgLeuLe 1467  
|||||  
4363 CTGTTGAGAGAGACACAGGAGCTGTGTGCCACCTGTGCTCAGGCTCT 4412  
1467 uArgHisCysGlySerArgIleSerThrIleArgThrHisAlaSerAlas 1484  
|||||  
4413 ACGACACTGTGGACGCCGATACGACCATCCGACGACAGCCAGCGCT 4462  
1484 eLeuTyrLeuLeuMetArgGlnAsnPheGluIleGlyHisAsnPheAla 1500  
|||||  
4463 CGCTGTACCTGCTCATGCGACAGAACTTGAGATCGGCGCAACAATTGGC 4512  
1501 ArgValLysMetGlnValThrMetSerLeuSerSerLeuValGlyThrTh 1517  
|||||  
4513 CGGTGAAAGATGAGGTCAACATGTCTCTGCTGCTCGTGGGAGCAGC 4562  
1517 rGlnAsnPheSerGluGluHisIleuArgArgSerLeuLysThrIleLeuT 1534  
|||||  
4563 GCGAACTTCAGTGAAGAGACACTGGGACGTTCACCTCAAAACCACTTCA 4612  
1534 hTyrAlaGluGluAspMetGlyLeuArgAspSerThrPheAlaGluGln 1550  
|||||  
4613 CCTATCTGAGGAGGACATGGGGCTGGGGAGACGACCTTCCAGAGCAG 4662  
1551 ValGlnAspLeuMetPheAsnLeuHisMetIleLeuThrAspThrValLy 1567  
|||||  
4663 GTCCAGGACCTGATGTTTCAACCTGCACATGATCTGCAGGACGCTGA 4712  
1567 sMetLysGluHisGlnGluAspProGluMetLeuIleAspLeuMetTyrA 1584  
|||||  
4713 GATGAGGAACCAACAGAGAGACCTGTGATGCTCATCGACTCAATGTACA 4762  
1584 rGluAlaArgGlyTyrGlnGlySerProAspLeuArgLeuThrTyrLeu 1600  
|||||  
4763 GAATTCGCCGGGGCTACAGGGCTCACCGGACCTTCGGCTGACCTGTGG 4812  
1601 GlnAsnMetAlaGlyLysHisAlaGluLeuGlyAsnHisAlaGluAlaAl 1617  
|||||  
4813 CAGAACATGCGCGGGAAGCAGCGGAGCTGGGCAACCAACGCGAGGCGCG 4862  
1617 aGlnCysMetValHisAlaAlaAlaLeuValAlaGluTyrLeuAlaLeuL 1634  
|||||  
4863 CCAGTGCATGTGTGACAGCGGCGCCCTGTGGCTGAGTACCTCGGCCCTGC 4912  
1634 euGluAspGlnArgHisLeuProValGlyCysValSerPheGlnAsnIle 1650  
|||||  
4913 TCGAGAGCACGCGCACCTGCGCGGTGGGTGCTTCCAGAACATC 4962  
1651 SerSerAsnValLeuGluGluSerAlaIleSerAspAspIleLeuSerPr 1667  
|||||  
4963 TCATTCACAGCTGTAGAGAGATCCGCCATCTCCGACGACATCTCTGCC 5012  
1667 oAspGluGluGlyPheCysSerGlyLysHisPheThrGluLeuGlyLeuV 1684  
|||||  
5013 CGAGCAGAGAGGGGCTTCTGCTCGGGAAGCACTTCACTGAGTGGGGCTGG 5062  
1684 aIGlyLeuLeuGluGlnAlaIlaGlyTyrPheThrMetGlyLysLeuTyr 1700  
|||||  
5063 TAGGGTTCGTGAAACAGGACCGGCTACTTCAACATGGCGGGCTTAC 5112  
1701 GluAlaValAsnGluValTyrLysAsnLeuIleProIleLeuGluAlaHis 1717



```
|||||
5113 GAGGCGGTGAATGAGTCTACAGAACCTCATCCCATCCGAGAACGCCA 5162
1777 SARGASPIRYTYSIYSLAVALAVALHISGLYLSLEUGLNUA1AP 1734
|||||
5163 CCGTGACTACAAAGAGAGTGGCCGGGTGCACGGCAAACTGCAGAGAGGCT 5212
1734 hethrLysIleMetHisGlnSerSerGlyTyrGluArgValPheGlyThr 1750
|||||
5213 TCACCAATCATGTCACAGAGATTCCGGCTGGAGCGGCTGTGGAGCG 5262
1751 TyrPheArgValGlyPheTyrGlyAlaHisPheGlyAspLeuAspGlu 1767
|||||
5263 TATTTCCGCGGTGCTTACGGGGCCACTTCGGTGACCTCGAATGAGCA 5312
1767 nGluPheValIlyIrysgluProSerIleThrLysLeuAlaGluIleSerH 1784
|||||
5313 GGAGTTGTGTACAGAGCATCATCAGAACCTGCAGAGATCTCAC 5362
1784 IARGLEUGLUGLUPHETHRGUAARGPHEGLYASPARVALVALGLU 1800
|||||
5363 ACCGGCTGAGAGATTCTACAGAGAGATTGGCGAGCGAGCTGTGAG 5412
1801 IleIleLysAspSerTyrProValAspLysSerLysLeuAspSerGlnly 1817
|||||
5413 AFTATCAAGACTCTTACCTCGTGGACAGTCCAAAGCTTGACTACAAA 5462
1817 SALATYRILEGLNILETHRTRYVALGLUPROTYRPHASPTHRTRYGLU 1834
|||||
5463 GGCTACTATCCAGATCAGTATGTGGAACCTACTTGATACCTACGAGC 5512
1834 eulysAspArgValThrTyrPheAspArgAsnTyrGlyLeuArgThrPhe 1850
|||||
5513 TCAAGAGACCGGCTACCTACTTTACCGCACTATGGGCTTCGCACATTC 5562
1851 LeuPheCysThrProPheThrProAspGlyArgAlaHisGlyLeuLeupr 1867
|||||
5563 CTGTTCTGCACAGCGCTTCACCGCCGATGGGGCGCACACGGGAGGTGCC 5612
1867 oGluGlnHisLysArgLysThrLeuLeuSerThrAspHisAlaPheProT 1884
|||||
5613 CGAGCAACACAAAGCGTAAGAGCGTCTCAGACACGACCCCTTCCCT 5662
1884 YRILELSTHARGLIeARVALYSHISARGGLUGLUTHRVALLEUTHR 1900
|||||
5663 ACATCAAGACTCGATCCGTTGTGCCACCGGAGGAGACGGTCTACG 5712
1901 ProValGluValAlaIleGluAspMetGlnLysLysThrArgGluLeuAl 1917
|||||
5713 CCAGTGGAGGTGGCCATCGAGAGCATGCAGAAAGACGGGAGGTGCC 5762
1917 aPheAlaThrGlnGlnAspProProAspAlaLysMetLeuGlnMetValL 1934
|||||
5763 CTTTGCCACCGAGAGAGACCCAGATGCTAAGATGCTACAGATGTGTC 5812
1934 euGlnGlySerValGlyProThrValAsnGlnGlyProLeuGluValAla 1950
|||||
5813 TTCAGGGGTCTGTAGGGCCACCGCTGAACAGAGGTCCCTCGAGAGTGGC 5862
1951 GlnValPheLeuAlaGluIleProGluAspProLysLeuPheArgHisH 1967
|||||
5863 CAGGTGTTTTTAGCAGAGATCCCGAAGACCCCAAGCTTCCGCGCATCA 5912
1967 sAsnLysLeuArgLeuCysPheLysAspPheCysLysLysCysGlnAsp 1984
|||||
5913 CAARAAATTCGGGCTGTGCTCAAGAACTTCTGAAAGAAATGTAGAGATG 5962
1984 IaleuArgLysAsnLysAlaLeuIleGlyProAspGlnLysGluTyrHis 2000
|||||
5963 CGCTGCGGAAATTAAGGCCCTGATTTGGGCCGAGCAAGAAAGAGTACAC 6012
2001 ArgGluLeuGluArgAsnTyrCysArgLeuArgGluAlaLeuGlnProLe 2017
|||||
```

```
6013 CGTGAGCTGAGGCGCAACTACTCCCGCTGGCGGAGGCTCTGAGCCCT 6062
2017 uLeuThrGlnArgLeuProGlnLeuMetAlaProThrProProGlyLeuA 2034
|||||
6063 GCTTACCAGCGCGCTGCCAGCTGATGGACACCACCCAGCGGCTCA 6112
2034 rGAsnSerLeuAsnArgAlaSerPheArgLysAlaAspLeu 2047
|||||
6113 GGAAGTCTTTGAAACAGAGCAAGTTTCCGAAAGGCAAGACCTC 6153
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/AA2001A.DAT:AAH41934
seq_documentation_block:
ID AAH41934 standard; cDNA: 6828 BP.
XX
XX AAH41934;
AC
XX
DT 30-AUG-2001 (first entry)
XX
DE Human CLASP-3 cDNA sequence SEQ ID NO:1.
XX
KW Human; cadherin-like asymmetry protein; CLASP; CLASP-3; immune response;
KW cell surface molecule; transmembrane protein; immunosuppressive; vaccine;
KW antiinflammatory; antiarthritic; antianaemic; dermatological; uropathic;
KW ophthalmological; antirheumatic; nephrotoxic; antithyroid; antidiabetic;
KW neuroprotective; antiscistmatic; antibacterial; antisense therapy;
KW gene therapy; chromosome 1p31.1; ss.
XX
OS Homo sapiens.
XX
PN WO200142297-A2.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-US34171.
XX
PR 13-DEC-1999; 99US-0170453.
XX
PR 14-JAN-2000; 2000US-0176195.
XX
PR 14-FEB-2000; 2000US-0182296.
XX
PR 11-APR-2000; 2000US-0196267.
XX
PR 11-APR-2000; 2000US-0196460.
XX
PR 11-APR-2000; 2000US-0196527.
XX
PR 11-APR-2000; 2000US-0196528.
XX
PR 11-APR-2000; 2000US-05472726.
XX
PR 13-OCT-2000; 2000US-0240503.
XX
PR 13-OCT-2000; 2000US-0240508.
XX
PA (ARBO-) ARBOR VITA CORP.
XX
PI Lu P, Gartman JD, Candia AF;
XX
DR WPI: 2001-375003/39.
XX
DR P-PSDB: AAB99495.
XX
PT Novel isolated cadherin-like asymmetry protein (CLASP) -3, useful for
PT useful in preventing or treating a CLASP-3-mediated disease e.g.
PT autoimmune disease
XX
XX Claim 1; Fig 6; 189pp; English.
XX
XX The present invention describes an isolated polypeptide (I) comprising a
XX nucleotide (nt) sequence (S1a) that has at least 90% identity to the
XX sequence given in AAH41934 and is immunologically cross-reactive with
XX the derived amino acid (aa) sequence (S1b) given in AAB99495 or shares a
XX biological function with native CLASP-3. (I) has immunosuppressive,
XX antiinflammatory, antiarthritic, antianaemic, dermatological, uropathic,
XX ophthalmological, antirheumatic, nephrotoxic, antithyroid, antidiabetic,
XX neuroprotective, antiscistmatic and antibacterial activities, and can be
XX used in antisense therapy, vaccine production and gene therapy. CLASP-3
XX related sequences can be used in preventing or treating a CLASP-3
XX mediated disease, preferably an autoimmune disease by inhibiting an
XX immune response. The autoimmune disease is caused or exacerbated by
XX increased activity of T helper 1 lymphocytes (TH1). AAH41893 to AAH41953
```

CC and AAB99491 to AAB99507 represent sequences which are used in the  
CC exemplification of the present invention. CLASP-3 is localised in the  
CC chromosome location 1p31.1.

xx Sequence 6828 BP; 2082 A; 1384 C; 1432 G; 1930 T; 0 other;

## alignment\_scores:

Quality: 7198.50 Length: 2140  
Ratio: 3.960 Gaps: 16  
Percent Similarity: 84.953 Percent Identity: 64.953

## alignment\_block:

US-09-736-968A-2 x AAH41934 ..

Align seg 1/1 to: AAH41934 from: 1 to: 6828

```
2  AAlaSerGluArgGlnAlaPheAlaHisLysIleAsnArgThrValAl 18
||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
19 GCCATGGCCGCGCGCGCGCTTCCGCCAGAAAGATCAGCAACGGTGGC 68
18 aAlaGluValArgLysGlnValSerArgGluArgSerGlySerProHis 35
||||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
69 AGCCGAAGTTAGAAAGCATCTCCGACAAATATAGTGTCTCCCAAC 118
35 eSerArg..... ArgCysSerSerSerLeuGlyVal 45
::: ::::::::::::::::::::::::::::::::::::::::::::::::::::
119 TGCACAAAACCTTAATATTTGTTGGCAATATATCCATCAACACAGTGTG 168
46 ProLeuThrGluValValGluProLeuAspPheGluAspValLeuLeuSe 62
||||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
169 CCCCTTACCGAAGCATAGATCCAGTGTGATTTGGAGATTAACCTCATTTAC 218
62 rArgProAspAlaGluProGlyProLeuArgAspLeuValGluPheP 79
::: ::::::::::::::::::::::::::::::::::::::::::::::::::::
219 TCATCTTTGGCTGTGATTTGGGCTTTACGGGATTTGATTAATTTTC 268
79 roAlaSerAspLeuGluLeuLeuLeuGlnProArgGluCysArgThrThr 95
||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
269 CTCAGATGATATGAAAGTTGTTATGTCCTCGGACTCGAAGACTCTT 318
96 GluProGlyLeuProLysAspGluLysLeuAspAlaGlnValAlaAla 112
::: ::::::::::::::::::::::::::::::::::::::::::::::::::::
319 GTTTCAGCTGTACCTGAAAGAAAGTAAATGATCCACATGTTAAAGACTG 368
112 aValGluMetThrIleGluAspTrpValIleValHisArgArgThrGlnT 129
::: ::::::::::::::::::::::::::::::::::::::::::::::::::::
369 TATAAGAAAGTTATACAGAAAGACTGGCAATTTGTATCAGAAAATATCATA 418
129 yrLeuSerAlaAlaLysSerProValThrThrAspThrGlnArgGluArg 145
||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
419 AATTGGGAACAGATTTAATCCCAATATAGTAAACAGAAAGAAAGG 468
146 GlnLysGlyLeuProArgGlnValPheGluGlnAspAlaSerGlyAspG 162
||||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
469 CAAAAGGTTTGGCCAAAACAAAGTTTGAATCTGATGACGCTCCAGATGG 518
162 uArgSerGlyProGluAspSerAsnAspSerArgArgGlySerGlySerP 179
::: ::::::::::::::::::::::::::::::::::::::::::::::::::::
519 CACAGCTACACAGATGATCAAGATGACCTTAAAGACGTTCAATGTCAA 568
179 roGluAspThrProArgSerSerGlyAlaSerSerIlePheAspLeuArg 195
::: ::::::::::::::::::::::::::::::::::::::::::::::::::::
569 TAGATGATACCCCAAGGGGTAGCTGGGCTGTATATCTTTGACTTTCAA 618
196 AsnLeuAlaAlaAspSerLeuLeuProSerLeuLeuGluArgAlaAlaPr 212
||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
619 AATTCACCTCTCTGATGCTTTGCTTCCCAATTTACTTATGATCAACTCCA 668
212 oGluAspValAspArgArgAsnGluThrLeuArgArgGlnHisArgProp 229
||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
669 TGAAGAAATTAAGCCGTGAGATGATGACCAAGAAATCAAAACCGTCA 718
```

```
229 roAlaLeuLeuThrLeuThrProAlaProAspGluAspGluAlaValAlu 245
||||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
719 AAGAACTTTTGTGCTTTCATCCATCACCAGATGAGAGAAACCAATTAACA 768
246 ArgCysSerArgProGluProArgGlnHisPheGlyGlnArgIleLe 262
||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
769 CGGCTTAGTGTCTCGATATACCAAGAAACATTTTGTGCAAGACTTCT 818
262 uValLysCysLeuSerSerLeuLysPheGluIleGluIleGluProIlePhe 279
||||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
819 TGTAAATCTTATCATCTCACTCAAGTTGAAATTTGAATTCACCATTTTGTG 868
279 yLLeuAlaLeuThrAspValArgGluLysLysIleSerGluAsn 295
::: ::::::::::::::::::::::::::::::::::::::::::::::::::::
869 CAAGTTGGCTTTATATGATGTCAAGAAAGAAAGAAAGATTTCCAGAAAC 918
296 PheThrPheAspLeuAsnSerAspSerMetLysGlyLeuLeuArgAlaH 312
||||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
919 TTTTATTTGACCTTATTTCTGACGATGAAAGGTTGTTACGTCCACA 968
312 sGlyThrHisProAlaIleSerThrLeuAlaArgSerAlaIlePheSer 329
||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
969 TGTACACACTGCTGCCATCTACTACCTGCGCAGATCAGCAATTTTCTCA 1018
329 aThrThrProSerProAspIlePheLeuValIleLysLeuGluLysVal 345
::: ::::::::::::::::::::::::::::::::::::::::::::::::::::
1019 TCACCTATCTCTCCCAAGATGTTTCTGTATTAAGCTTACGAAAGATC 1068
346 LeuGlnGlnGlyAspIleSerGluCysGluProTyrMetValLeuL 362
||||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
1069 CTACACCAAGAGACATTTGGAGAGTGTGCAGAACCATATATGATTTTCAA 1118
362 sGluValAspThrAlaLysAsnLysGluLysLeuGluLysLeuArgLeu 379
||||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
1119 AGAAGCAGATGCCACCAAGATTAAGAAACTGGAGAAATCGAAGAGTTC 1168
379 laAlaGluGlnPheCysThrArgLeuGlyArgTyrArgMetProPheAla 395
||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
1169 AAGCAGATCAGTTTGTCCAAAGACTGGGAAATATCGCTGCTTTTCTT 1218
396 TrpThrAlaValHisLeuAlaAsnIleValSerSerAlaGluGlnLeuAs 412
||||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
1219 TGGACTGCAATCCATTTTAAATGAATTTGTTACAGAGTGTGGAGAGTTTGA 1268
412 pArgAspSerAsp..... SerGluGlyGluArgArgProAlaT 425
::: ::::::::::::::::::::::::::::::::::::::::::::::::::::
1269 AAGAGTTTCTACAGAAAGTAAATCAGTACTGAGACAGAAAGGCTCTT 1318
425 rThrAspArgArg..... ArgGlyProGlnAspArg 436
||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
1319 GGTCAAGAGAGAGAAATTTAGTATTTGTCGACAGATCACTTGAAGG 1368
437 AlaSerSerGlyAspAspAlaCysSerPheSerGlyPheArgProAlaTh 453
::: ::::::::::::::::::::::::::::::::::::::::::::::::::::
1369 ACACACAGAGGAGATGATCTGTGAATCTTGACAGAGCTTTCGACAGCTAC 1418
453 rLeuThrValThrAsnPhePheLysGlnGluAlaGluArgLeuSerAspG 470
||||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
1419 TCTCAGAGTGACAAATTTTAAAGCAGAGAAAGACCGCTTAATGTATG 1468
1469 AAGATCTTACAAATTCCTTGTGATATGAGAAGGCCATCTTCTGTCTTA 1518
470 LuAspLeuPheLysPheLeuAlaAspMetArgArgProSerSerLeuLeu 486
||||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
1519 CGGCGCTAAGACCTATTACACTCAGCTCAAGATAGACATTTCTCCGC 1568
487 ArgArgLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAl 503
||||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
503 rProGluAsnProHisPheCysLeuSerProGluLeuLeuHisIleLysP 520
||||| ::::::::::::::::::::::::::::::::::::::::::::::::::::
1569 ACCTGAAATTCGCCATATTTGCTTAAGTCCGAGAGCTGTTCAAGTGAAC 1618
520 rThrProAspProArgGlyArgProThrLysGluIleLeuGluPhePro 536
```

```
||||||| ||| |||||||:|||||||||
1619 TTACCCCTCAGTAGAGTTAGACCTACAGAAATCTTAGAGTTTCCC 1668
1619 AlaArgGluValIlePheAlaProHisThrSerIlePheArgAsnLeuLeuTyrVal 553
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1669 GCAGGGAGATTGATTGTTCCAAACACTACTTACAGAAATCTTCTCTACAT 1718
553 LysProHisSerLeuAsnPheSerSerArgGlnIleSerValIleArgAsnL 570
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1719 ATACCTCAGAGCTTAAATTTTGGCAATCGTCAAGTCTCTAGAAATA 1768
570 euAlaValArgValGlnTyrMetThrGlyIleAspProSerGluAlaLeu 586
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1769 TACAGGTGAAGTCCAGTTTATGTATGGAGAGATCCAGCAAGTCCCAAG 1818
587 ProValIlePheGlyLysSerSerCysSerGluPheThrArgGluAlaPhe 603
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1819 CCGGTAACTTTGGTAAATCTAGCTGTTACGAATTTTCAAGGAAGCCTA 1868
603 eThrProValValTyrHisAsnLysSerProGluPheTyrGluIlePheL 620
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1869 TACAGCCGTATATCATTAACAGGCTCTCTGATTTTCATGAAAGAAATCA 1918
620 LysLeuHisLeuProAlaCysValThrGluAsnHisLeuLeuPheThr 636
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1919 AGGTAAAGCTCTCTGCTACTTTAACTGACATCATCTGCTGCTTTTACT 1968
637 PheTyrHisValSerCysGlnProArgProGlyThrAlaLeuGluThrPr 653
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1969 TTTTATCATGTTAGTTGTCTACAAACAAACAAATACTCTCTTGAAACACC 2018
653 oValGlyPheThrTyrIleProLeuLeuGlnHisGlyArgLeuArgThrG 670
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2019 AGTTGATATACATGATGATACCATGCTTCCAGAAATGAGCGTTGAAGACTG 2068
670 LysProPheCysLeuProValSerValAspGlnProProSerSerIleSer 686
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2069 GCCAGTTTGTCTGCCAGTCTCATTTGGAAACACCCACAGGCTTATTTCT 2118
687 ValLeuThrProAspValAlaLeuProGlyMetArgTyrValAspGlyH 703
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2119 GTACTGTCTCTCAGGTCCTGCTCACTGCGCATGAATGGGTAGATAATCA 2168
703 sLysGlyValAlaPheSerValGluLeuThrValAlaSerSerValHisProG 720
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2169 CAAGAGTGTTTATGTTGAAGTGTGCTGTTCGTATCTCCATACAC 2218
720 LysProTyrLeuAspLysPhePheThrLeuValHisValLeuGluGlu 736
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2219 AAGATCCTTATCTTGACAAATTTTGTGCTGTCATGCTCTGGATGAA 2268
737 GlyAlaPheProPheArgLeuLysAspThrValLeuSerGluGlyAsnVal 753
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2269 CACCTGTCTCCAGTCCGAATTTGGGACATGCAATCATGGAAATAACTT 2318
753 LglGlnGluLeuArgAlaSerLeuAlaAlaLeuArgLeuAlaSerProG 770
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2319 AGAAATGAAATGAAAGACAGATATTTCAGCACTGATCATCCACAGCTGG 2368
770 LysProLeuValAlaPheSerHisHisValLeuAspLysLeuValArgLeu 786
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2369 AACCAATGTGGCGATTTCTTCACTTCTGCTAGATTAACCTGATACTTTTA 2418
787 ValIleArgProProIleIleSerGlyGlnIleValAsnLeuGlyArgG 803
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2419 GTATATTAGACCTCTCTCATGCTGGCCAAATAGTTAACTGATGCTCAAGC 2468
803 yAlaIlePheGluAlaMetAlaHisValAlaSerLeuValHisArgSerLeuG 820
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2469 ATCTTTTGAAGCATGCAATCAATTAAATGCACTTCACAAAACCTTGG 2518
820 LuAlaIleGlnAspValArgGlyHisCysProGlnLeuAlaIleTyrVal 836
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
2519 AAGAAATCATGACGACATGGCAGAAACAGCCTTCTTGATCATATATT 2568
837 HisTyrAlaPheArgLeuProGlyIleThrGluProSerLeuProAspGlyAl 853
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2569 CATTTATGTTTCCGCTACCAAAATACTTACCTTAATCATCATGACACAGG 2618
853 aPro.....ProValThrValGlnAlaAlaThrLeuAlaArgGlyS 867
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2619 TCTGTGGGGTTTGGAGAGATGATGCTATATGCAATATGCTGATGATCTG 2668
867 eGlyArgProAlaSerLeuTyrLeuAlaArgSerLysSerIleSerSer 883
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2669 CCGTGAAGACCTGCAGACCTTATTTAAATCGTTCTCGAAGCCTTAGTAAT 2718
884 SerAsnProAspLeuAlaValAlaProGlySerValAspAspGlyValSe 900
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2719 AGCAATCCAGATATATCTGGGACTCCACAGTCACAGATGTAAGTTGG 2768
900 rArgIleuAlaSer..... 905
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2769 ATCAATCATCGGGAGTAAAGGCTTTAGATCGCTCCCAATTCTCGGTTAACA 2818
905 ..... 905
2819 CTGTGTGTCCAAAAGCTGCCCATGGGAGTCCAAACCCAGTCCAAAGTGA 2868
905 ..... 905
2869 GATCAACACAGGCTATGATGATGCAAGTTGTATGATGTCTTGCAACAC 2918
906 .....L 906
2919 AGAGACGTCAAGTTTCTTACAAACATTAAACGGACGCTTACCAACTAAA 2968
906 LysLeuHisGluGluLeuAlaLeuGlnTyrValValSerSerSerAla 922
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2969 AGCTTTTACAGAGAGAGCTGGCTTTCAGATGGGTGTTTGGAGTGGACGC 3018
923 ValArgGluAlaIleLeuGlnHisAlaThrPhePhePheGlnLeuMetVal 939
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3019 GTTCGGGAATCAGCTTTCACAGAACGTTTCATGATGATGACATGCAAGCTTGTGC 3068
939 LysSerMetAlaLeuHisLeuLeuLeuGlyGlnArgLeuAspThrProA 956
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3069 AAAGACATAGTGCACCATTTTACTTTAATGATATAACTTAGAGCTCCAA 3118
956 rGlyLysLeuArgPheProGlyArgPheLeuAspAspIleThrAlaLeuVal 972
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3119 GCAAAAGTGTGTTTCCAGAACGTTTCATGATGATGACATGCAAGCTTGTGC 3168
973 GlySerValGlyLeuGluValIleThrArgValHisLysAspValGluLe 989
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3169 AGCACATTTGCTAGTATATATGTTTCACAGATTTTCAGAAAGACACAGAAAT 3218
989 uAlaGlnHisLeuAsnAlaSerLeuAlaPhePheLeuSerAspLeuLeuS 1006
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3219 GATTGAGAACATCAATACAAACCTTGATCTTCTCATATATCTGTTGT 3268
1006 eTyrLeuValAspArgGlyPheValAlaPheSerLeuValAlaGlnHisTyrLys 1022
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3269 CTGTATGGACAGAGATTTGTTTATAGCTTATAAAGTCCGCTATATAA 3318
1023 GlnValAlaThrArgLeuGlnSerSerProAsnProAlaAlaLeuLeuTh 1039
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3319 CAGGTGCTTCAAGAGCTTACTCATTTACAGAAATCCAGATGTTGTGTTGTC 3368
1039 rLeuArgMetGluPheThrArgIleLeuCysSerHisGlnHisTyrValIT 1056
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3369 CTTGAGGCTGGATTTTCTACGAATCATCTGCACTATAGACATATGTGTA 3418
1056 hTyrLeuAsnLeuProCysCysProLeuSerProProAlaSerProSerPro 1072
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3419 CATTAACCTTACCCCTGACACTTACTTACTTCCACCTGCATCTCCATACCT 3468
```

```
1073 SerValSerSerThrThrSerGlnSerSerThrPheSerSerGlnAlaIpr 1089
|||||
3469 TCTGTTCTTGTGCAACATCTCAGAGTTCGTGATTTCTTACGAATGTGCA 3518
|||||
1089 GAsProIysValLThrSerMetPheGlnLeuSerGlyProPheArgGln 1106
|||||
3519 ACACCAAAAGATTGCCAATATGTGTAATTTATCCGTGCTTCCGCCAAC 3568
|||||
1106 LnhIsPheLeuAlaGlyLeuLeuLeuThrGluLeuAlaLeuAlaLeuGlu 1122
|||||
3569 ACCATTATTGGCAGGACTTGTGTTAACAGAGCTGCTGTCATTPTTAC 3618
|||||
1123 ProGluAlaGluGlyAlaPheLeuLeuHisLysLysAlaIleSerAlaVa 1139
|||||
3619 CCTGATGCTGGAAGACGTGTTGGATTGCATGAAGAAGATCAATCAATGT 3668
|||||
1139 LhIsSerLeuLeuCysGlyLhIsAspThrAspProArgTyraIleGluAla 1156
|||||
3669 ACACAAATTACTCTCCAGTCAGACTCAGACCCGGGTACTGTGACCTTC 3718
|||||
1156 hVAlLysAlaArgValAlaGluLeuTyrrLeuProLeuLeuSerIleAla 1172
|||||
3719 AGATAAGGCTCGAGTGGCCATGTGTATCTACTCTGATGGTATTTATWC 3768
|||||
1173 ArgAspThrLeuProArgLeuHisAspPheAlaGluGlyProGlyGlnAr 1189
|||||
3769 ATGGAACAGTACCTCAGCTGATGATGTTTACAGAACTCACAAATCAACG 3818
|||||
1189 gSerArgLeuAlaSerMetLeuAspSerAspThrGluGluGluAspI 1206
|||||
3819 AGGAACACCAATTTGTATAGCCATGATGATATGAAGTAGAGC.... 3864
|||||
1206 IeAlaGlyThrIleAsnProSerValAlaMetAlaIleAlaGlyPro 1222
|||||
3865 ..GGAAGTATGATMAACCCAGACCGTGGCCATGCGCAATCCAGGAGCATCG 3912
|||||
1223 Leu.....AlaProGlySerArgAlaSerIleSerGlnGlyPr 1235
|||||
3913 GTCCCTCACTAACAGCGCTGCGACTTCCCTCAGCTCA..... 3954
|||||
1235 oProThrAlaSerArgAlaGlyCysAlaLeuSerAlaGluSerSerArgT 1252
|||||
3955 ....ACGAGTGGCAGCAACACACTACTTTCACGACAGATCAAGTCGAA 4000
|||||
1252 hTrLeuLeuAlaCysValLeuTrpValLeuLysAspThrGluProAlaLeu 1268
|||||
4001 GCCTTTGATCTGTCTACTTGGTCTCAAAATGCAAGTCAACAGATT 4050
|||||
1269 LeuGlnArgTrpAlaThrAspLeuThrLeuProGlnLeuGlyArgLeuLe 1285
|||||
4051 CTACAGAGAGTGTACAGATCTCAGTCTTCAGCTTAACCGGCTATT 4100
|||||
1285 uAspLeuLeuTyrrLeuCysLeuAlaAlaPheGluTyrrLysGlyLysLysA 1302
|||||
4101 AGATCTGCTTATCTGTGTCTGCTTGGCTTGAATAAGGAAAAAAG 4150
|||||
1302 LAbheGluArgGlyIleAsnSerLeuThrPheLysLysSerLeuAspMetLys 1318
|||||
4151 TGTGTAACCAATGATAGCTTACCTTTTAAAGAAATCAAAAGCATGAGA 4200
|||||
1319 AlaArgLeuGluAlaIleLeuGlyThrIleGlyAlaArgGlnGluLe 1335
|||||
4201 GCAAAAGCTTGAAGAGCTATTCTGGAGCATAGTGCCAGCAAGAAAT 4250
|||||
1335 tValArgArgSerArg.....GluArgSerPro.....P 1345
|||||
4251 GGTACGGCGAAGCCGAGACAGCTCGAGAGAAACCCATCTGGAAGTGCC 4300
|||||
1345 hGlyAsnProGluAsnValArgTrpArgLysSerValThrHisTrpLys 1361
|||||
4301 TTGGAAATCAAGAAATTTGAGGTGGAGGAAGAAATATGACTCTGGCCT 4350
|||||

1362 GlnThrSerAspArgValAspLysThrLysAspGluMetGlnHisGluAl 1378
|||||
4351 CAAAACACAGAGAAGCTGTGCAAAATCAAGAGCAGAGATTGAACACGAA 4400
|||||
1378 aLeuValGluGlyAsnLeuAlaThrGluAlaSerLeuValValLeuAspT 1395
|||||
4401 ACTGATTGATGAGAACTGGCTACAGAGCAACCAATCAATCATTTTAAATA 4450
|||||
1395 hTrLeuGluIleIleValGlnThrValMetLeuSerGlnAlaArgGluSer 1411
|||||
4451 CATTAGAGATTGTGTTCCAGACCGCTTCTGTACGGAATCCAAAGAGAGC 4500
|||||
1412 ValLeuGlyAlaValLeuLysValValLeuTyrrSerLeuGlySerAlaG 1428
|||||
4501 ATTCTGTGTGAGTGTAAAGTCTACACAGATGCGCTGTAAACA 4550
|||||
1428 nSerAlaLeuPheLeuGlnHisGlyLeuAlaThrGlnArgAlaLeuValS 1445
|||||
4551 AAGTGCAGTTTATCTACACACGTGTTTGTCTACACAGAGAGCCCTGGTTT 4600
|||||
1445 eTrLysPheProGluLeuLeuPheGluGluLysAspThrGluLeuCysAlaAsp 1461
|||||
4601 CAAGCTTCTCGAAGCTCTTATTGGAACAGACAGACAGAGCGTGTGCTGAT 4650
|||||
1462 LeuCysLeuArgLeuLeuLeuArgHisCysGlySerArgIleSerThrIleAr 1478
|||||
4651 TTATGCCCTCAGGCTTCCGACACGTGACAGATGCGTCAATATACG 4700
|||||
1478 gThrHisAlaSerAlaSerLeuTyrrLeuLeuMetArgGlnAspPheGlu 1495
|||||
4701 GTCACACCCCGAGTCCCTCCCTTACTACTATGAGGCAAAACTTTGAGA 4750
|||||
1495 IeGlyHisAsnPheAlaArgValLysMetGlnValThrMetSerLeuSer 1511
|||||
4751 TTGGGAATTAATTTGCCAGGTTAAATGCAGGTACCAATGTCACATATCC 4800
|||||
1512 SerLeuValGlyThrThrGlnAsnPheSerGluGlnHisLeuArgArgSe 1528
|||||
4801 TCTTGTGGTGGACATCTCAGAAATTTTAAATGAATAATTTTAAAGACGTTC 4850
|||||
1528 rLeuLysThrIleLeuThrTyrrAlaGluGluAspMetGlyLeuArgAspS 1545
|||||
4851 TCTAAAGCTATATTGACATATGCTGAAGAAAGATCTGGAATGGAGGAAA 4900
|||||
1545 eTrThrPheAlaGluGlnValGlnAspLeuMetPheAsnLeuHisMetIle 1561
|||||
4901 CAACATTTCTGATCAGATCAGGTCAGGATCTGTGTTCAATCTCCATATGATT 4950
|||||
1562 LeuThrAspThrValLysMetLysGluHisGlnGluAspProGluMetLe 1578
|||||
4951 CTTCTGTACTGTGAATAATGAAGACACACCGAGAGATCTGAAATGTT 5000
|||||
1578 uIleAspLeuMetTyrrArgIleAlaArgGlyTyrrGlnGlySerProAspL 1595
|||||
5001 GATTGATCTAATGTACACAATGGCCAAAGGTTTACACAGCTCTCCAGAG 5049
|||||
1595 euArgLeuThrTrpLeuGlnAsnMetAlaGlyLysHisAlaGluLeuGly 1611
|||||
5050 ..CGATTGACCTGTGTCAGAACATGCGCAGGCAAGCATCTCAAGACGAA 5097
|||||
1612 AsnHisAlaGluAlaAlaGlnCysMetValHisAlaAlaAlaLeuValAl 1628
|||||
5098 AATCATGTCTGAAGCTGCGACATGTCTTACTCCATCAGCAGACACTTGTTC 5147
|||||
1628 aGluTyrrLeuAlaLeuLeuGluAspGlnArgHisLeuProValGlyCysV 1645
|||||
5148 TGAATATTTGAGCATGCTGAGAGCCGGAATAATCTCTCTGGGAGTGTG 5197
|||||
1645 aIleSerPheGlnAsnIleSerSerAsnValLeuGluGluSerAlaIleSer 1661
|||||
5198 TAACATTTTCAGAAATTTTCAATCTAATGTGTTTAAAGAAATTCGGCGTCA 5247
|||||
1662 AspAspIleLeuSerProAspGluGluGlyPheCysSerGlyLysHisAsp 1678
|||||
```

```

|||||.....|
5248 GATGATGGGATCTCCAGATGAAAGAGTATCTGCTCTGGAATAACTCTT 5297
1678 eThrGluLeuGlyLeuValGlyLeuLeuGluGluAlaAlaGlyTyrPhe 1695
5298 TACTGAGTCAGACCTGTGGATTAAGTGAACAAGACAGCTCTCTCTCT 5347
1695 hMetGlyGlyLeuTyrGluAlaValAsnGluValTyrLysAsnLeuIle 1711
5348 CTATGGCTGGCATGTATGAACAGTATTAAGTTTACAAAGTCTTATT 5397
1712 ProIleuGluAlaHisArgAspTyrLysLysLeuAlaValHisGly 1728
5398 CCTATTATAGAGCTATTCGGAGTGAACAAGAACTATCCCAATTCTAGG 5447
1728 yLysLeuGluGluAlaPheThrLysIleMetHisGlnSerSerGlyTrp 1745
5448 TAACTTCAGAAAGCATTCAGCAAAATTTGTCATCAGATCTAGTGGTGG 5497
1745 LuArgValPheGlyThrTyrPheArgValGlyPheTyrGlyAlaHisPhe 1761
5498 AGCGGATGTTGGCACTATTTCGTGTGTGTTTATGGAACCAAGTTC 5547
1762 GLyAspLeuAspGluGluGluPheValTyrLysGluProSerIleThrLy 1778
5548 GGGGATTTGGATGAACAAGATTGTTTACAGAGAGCTGCATTAACCA 5597
1778 sLeuAlaGluIleSerHisArgLeuGluGluPheTyrThrGluArgPheG 1795
5598 ACTTGCAAGATATCTCACAGATTGGAGGATTTTACGGAAGAAATTTG 5647
1795 LyAspAspValAlaGluIleIleLysAspSerTyrProValAspLysSer 1811
5648 GAGAGGATGTGGTGAAGTATCAAAAGCTCTAATCTGTAGACAAAGTGT 5697
1812 LysLeuAspSerGlnLysAlaTyrIleGlnIleThrTyrValGluProTy 1828
5698 AAATTAATCTTAAACAAGCATATATCAAGTATACATATGAGGACATA 5747
1828 rPheAspThrTyrGluLeuLysAspArgValThrTyrPheAspArgAsn 1845
5748 CTTTGACACATATGAGATGAGAAAGCAAGATCACCTATTTCGACAAAAAT 5797
1845 yrcGlyLeuArgThrPheLeuPheCysThrProPheThrProAspGlyArg 1861
5798 ACAATCTTCGTGATTCATGTAACACCTTTACTTACTTACTTACTTACT 5847
1862 AlaHisGlyGlyLeuProGluGlnHisLysArgLysThrLeuLeuSerTh 1878
5848 GCCCATGGGGAAGCTTCATGAACAATTCAAAAGGAAGACCATTCGTACT 5897
1878 rAspHisAlaPheProTyrIleLysThrArgIleArgValCysHisArg 1895
5898 GTCTCATGCTTCTCTTATATTAAACAAGGTCATATGTCACCTATTAAG 5947
1895 LuGluThrValLeuThrProValGluValAlaIleGluAspMetGlnLys 1911
5948 AAGAGATCATCTTAACACCAATGAAGTTGCTATTGAGGACATGAGAAA 5997
1912 LysThrArgGluLeuAlaPheAlaThrGluGlnAspProProAspAlaLy 1928
5998 AAGACACAGAGAGTTGSCATTTGCACACATCAGAGATCCGACAGACCCA 6047
1928 sMetLeuGlnMetValLeuGluGlnLysSerValGlyProThrValAsnGln 1945
6048 AATGCTTCAGATGTACTCCAGGATCTGAGGACCAACAGTGAATCAGG 6097
1945 LyProLeuGluValAlaGlnValPheLeuAlaGluIleProGluAspPro 1961
6098 GGCTTTGGAGAGTTGCCAGGTTTCTGTGAAATACCTAGTGACCA 6147
1962 LysLeuPheArgHisHisLysLysLeuArgLeuCysPheLysAspPheCy 1978
|||||.....|

```

```

6148 AACCTCTCAGACATCATATAAATACTGCAGCTCTGCTTAAAGATTTCAC 6197
1978 sLysLysCysGluAspAlaLeuArgLysAsnLysAlaLeuIleGlyPro 1995
6198 TAAAGAGTGTGAGATGCTTTAAGAAAATAAGACCTTAATTTGGCCGG 6247
1995 spGlnLysGluTyrHisArgGluLeuGluArgAsnTyrCysArgLeuAr 2011
6248 TTCAAAAGAGGTATCAAGGAATTC...GGGAAGTATCTTCGCCCTTAA 6294
2011 gGluAlaLeuGlnProLeuLeuThrGlnArgLeuProGlnLeuMetAla 2027
6295 AGAGGCCCTACAGCCCT...AGATCACAGAAAGTCCCTCAGTTATCCAA 6341
2028 ProThr.....ProProGlyLeuArgAsnSerLeuAsnArgAlaSerPh 2042
6342 CCGATTTCTGCTTCTCCCTCCGACAGAGATCTCTTCTGAGTCGATAGCT 6391
2042 eArgLysAlaAspLeu 2047
6392 TCGCAAAATGATCTC 6407

```

seq\_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA507373

seq\_documentation\_block:

ID AA507373 standard; CDNA: 7215 BP.

AA507373:

26-SEP-2001 (first entry)

Human cDNA encoding CLASP-5.

Human; CLASP-5; cadherin-like asymmetry protein; immune gateway;  
immunogen; antibody; autoimmune disease; rheumatoid arthritis;  
multiple sclerosis; leukaemia; insulin dependent diabetes mellitus;  
acquired immunodeficiency syndrome; AIDS; ss.

Homo sapiens.

Location/Qualifiers

112..6048

/\*tag- a

/product= "CLASP-5"

1498..1519

/\*tag- b

/note= "PCR primer HC5gs1"

3087..3113

/\*tag- c

/note= "Antisense oligonucleotide AA507384"

3087..3113

/\*tag- d

/note= "Antisense oligonucleotide AA507384"

3232..3259

/\*tag- e

/note= "Antisense oligonucleotide AA507384"

complement (3642..3660)

/\*tag- f

/note= "PCR primer HC5As10b"

4728..4756

/\*tag- g

/note= "Antisense oligonucleotide AA507384"

WO200142296-A2.

14-JUN-2001.

13-DEC-2000; 2000MO-US34163.

13-DEC-1999; 99US-0170453.

14-JAN-2000; 2000US-0176195.

14-FEB-2000; 2000US-0182296.

11-APR-2000; 2000US-0196267.

```
PR 11-APR-2000; 2000US-0196460.
PR 11-APR-2000; 2000US-0196527.
PR 11-APR-2000; 2000US-0196528.
PR 11-APR-2000; 2000US-0547276.
PR 13-OCT-2000; 2000US-0240503.
PR 13-OCT-2000; 2000US-0240508.
PR 13-OCT-2000; 2000US-0240543.
PR 13-OCT-2000; 2000US-0240539.
XX (ARBO-) ARBOR VITA CORP.
XX
XX Lu P, Garman JD, Candia AF;
XX WPI: 2001-367865/38.
XX P-PSDB: AAU04024.
XX
XX CLASP-5 polynucleotides, proteins and antibodies are used to prevent or
XX treat a CLASP-5 mediated disease, such as an autoimmune disease e.g.
XX Rheumatoid arthritis -
XX
XX Claim 1: Fig 6; 188pp; English.
XX
XX The sequence encodes a cadherin-like asymmetry protein, CLASP-5, which
XX is a transmembrane protein of the immune system involved in the formation
XX of the immune gateway. CLASP-5, polynucleotides encoding it and an
XX anti-CLASP-5 antibody are used to prevent or treat a CLASP-5 mediated
XX disease, such as an autoimmune disease caused or exacerbated by
XX increased activity of TH1 cells. These diseases may include Addison's
XX disease, rheumatoid arthritis, dermatitis, Multiple Sclerosis, Systemic
XX Lupus Erythematosus and Autoimmune Thyroiditis, Inflammatory conditions
XX (e.g. Ischemia-reperfusion) and responses, Leukaemia, acquired
XX Immunodeficiency syndrome, AIDS, syphilis, typhoid, Burkitt's lymphoma,
XX Grave's disease and insulin dependent diabetes mellitus. CLASP-5 is used
XX for inhibiting an immune response in a cell (T cell or B cell) by
XX interfering with the expression of a CLASP-5 gene in the cell, the
XX ability of a CLASP-5 protein to bind to another cell or the ability of a
XX CLASP-5 protein to bind to another protein. The antibody binding CLASP-5
XX is used to inhibit an immune response in a subject. The polynucleotide is
XX used to detect CLASP-5 expression in cells and for diagnosis of
XX diseases and disorders associated with aberrant expression of CLASP-5.
XX
XX Sequence 7215 BP; 2005 A; 1734 C; 1721 G; 1755 T; 0 other;
XX
alignement_scores:
    Quality: 5258.50      Length: 2075
    Ratio: 3.299          Gaps: 24
    Percent Similarity: 76.819      Percent Identity: 51.614
alignement_block:
us-09-736-968a-2 x MAS07373      ..
Align seg 1/1 to: MAS07373 from: 1 to: 7215
36 SerArgArgcysSerSerSerleuGlyValPro...LeuThrGluValVala 51
   ::::::::::: ::::::::::: :::::
34 ACGGCGCGTGTACTGTGAATCTTTCAAACGCTCACTTTATGACCTGTg 83
51 IGLuProLeuAspPheGluAspValLeuLeuSerArgProProAspAlag 68
   |||||:::|||||::: ||:::
84 GGAACCGGTGACTTTGAAGACTTGTGATGACACACACCTGACACGCTGg 133
68 LuProGlyProLeuArgAspLeuValGluPheProAlaAspAspLeuGlu 84
   ::::::::::: ::::: ||::: |||||:::
134 ATGTGAGAGCTTGGCCAGAGAGCTGGGAGACTTCAGTGTGACGACTTGgAC 183
85 LeuLeuLeuGlnProArgGluCysArgThrThrGluProGlyLeuProGly 101
   ::::::::::: ||:::|||||::: ||:::
184 GTGGTGTTCACGCCAAAGAAATGTAGAGACTTTCAGCCCTCTTGGCGGA 233
101 sasp...GluLysLeuAspAlaGlnValArgAlaAlaValGluMetTyrI 117
   ::::: ||::: |||||::: ||:::
234 GGAAGGGGTTGAACTGGAACCTCATGTCAAGGAGACTGTGTTCAGACTTACA 283
```

```
117 IeGluAspTrpValIleValHisArgArgTyrGlnTyrLeuSerAlaAla 133
   ||:::|||||::: ||:::
284 TCCGTGAGTGGCTATGCTGTAACCGGAAACCAAGAAAGTCCAGAAATC 333
134 TyrSerProValThrThrAspThrGlnArgGluArgGlnLysGlyLeuPr 150
   ::::: ||::: ||::: |||||
334 TGTGGCTTTAAAGACTGATGATCGAAAGATTTTCACAAAGACGGCTTCC 383
150 cArgGlnValPheGluGlnAsp.....AlaSerGlyAspGluAla 163
   ||:::|||||::: ||:::
384 GAAACAGACGTTTGAATCGGAACCTTGAGTGCAGTAACCCGCTGCTC 433
163 rGserGlyProGluAspSerAsnAspSerArgArgGlySerPro 179
   ||:::|||||::: ||:::
434 AGCGAGCGCCCGCCGACTTAAGCTGCTGCGCAGCGTCTGGGAAAGGC 483
180 GluAspThrProArgSerSerGlyAlaSerSerIlePheAspLeuArg 196
   ||::: ||::: |||||::: ||:::
484 CCCGTCAC.....GCCGTGACTTTGACCTCCGACG 515
196 nLeuAlaAlaAspSerLeuLeuProSerLeuGluArgAlaAlaProG 213
   ||::: ||::: ||::: |||||::: ||:::
516 CCGCAGCCCTGACAAAGCGGCTAGAAACCTCTCAGCAAGTACGTCGC 565
213 LuAspValAspArgArgAsnGluThrLeuArgArgGlnHisArgProPro 229
   ||||| ::::: |||||
566 AGGACTTTGACAGACAGAAACGAGAGCGCCGAGACCAACAGCAGCGCC 615
220 AlaLeuLeuThrLeuTyrProAlaProAspGluAspGluValGluArg 246
   ||||| ||::: ||::: |||||::: |||||
616 GAGCTCTTGGCTTTTACCCTGAGTGCAGGAGGAGGAGGAGTGTGGAAAT 665
246 gCysSerArgProGluProProArgGluHisPheGlyGlnArgIleLeu 263
   ||||| ||::: ||::: |||||::: |||||
666 AGCTGACATACAGAGATGTCCCAAGAAACCTGGGCAACGAAATATTGG 715
263 aLysGlyCysLeuSerLeuLysPheGluIleGluIleGluProIlePheGly 279
   ||::: ||::: ||::: |||||::: ||::: ||:::
716 TCMACTGCTGACCTGACCTGAGTTCGAGATGAATTGACCCCTTTGGCC 765
280 IleLeuAlaLeuTyrAspValArgGluLysLysLysIleSerGluAsnPh 296
   ::::::::::: ||::: ||::: |||||::: |||||
766 AGCATTTGCCCTCTACGATGTTAAAGAAAGAAAGAAAGATCTGAGAAATTT 815
296 eTyrPheAspLeuAsnSerAspSerMetLysGlyLeuLeuArgAlaHisG 313
   ||::: |||||::: ||::: |||||::: ||::: ||:::
816 TCACGTGACCTGAACTCTGACACAGTTTCAAGATTTCTCGAGCTCACA 865
313 LyThrHisProAlaIleSerThrLeuAlaArgSerAlaIlePheSerVal 329
   ||::: ||::: ||::: |||||::: |||||
866 CGCCTTCAGTGGCCGATCAAGTCAAGCGGAGATCTGAGTCTTCAGTC 915
330 ThrTyrProSerProAspIlePheLeuValIleLysLeuGluLysValle 346
   |||||::: ||::: ||::: |||||::: |||||
916 ACCTACCCGCTCTCAGACATCTACTGTTAGTCAAGATGAAAAAGTCTC 965
346 uGlnGlnGlnAspIleSerGluCysGluProTyrMetValLeuLysG 363
   |||||::: ||::: ||::: |||||::: ||::: ||:::
966 GCAGCAGGAGATATTGGAGACTGTGCAGAGCCCTACAGGTTATCAAG 1015
363 LuValAspThrAlaLysAsnLysGluLysLeuGluLysLeuArgLeuAla 379
   ||::: ||::: ||::: |||||::: |||||
1016 AAAGTGAATGGTGAAGAGTAAGAAAGAAATTTGAAAAACTAAACTCAA 1065
380 AlaGluGlnPheCysThrArgLeuGlyArgTyrArgMetProPheAlaTr 396
   |||||::: ||::: ||::: |||||::: |||||
1066 GCTGAATCTTCTGCGACGCTTGGGAAATACCGGATGCCCTTGGCTG 1115
396 pThrAlaValHisLeuAlaAsnIle.....ValSerSerAlaGlyGlnL 411
   ||::: ::::: ||::: |||||::: ||:::
1116 GGCACCCATTAAGCTTATCAAGCTTCTCAATGCTCCACCCCTTGAGAGGG 1165
```

```

411 euaspargaspseraspser.....GluGlyGluArg 421
1166 AGGTAAGTGTGAGTGTGTGTTGGAGAAAGCCAGTGGGTGAGACGG 1215
422 ArgProAlaThrPheHisParGArgArgGlyProGlnAspArgAlaSe 438
1216 AGGACATGGCCCAATCTAGAAAGCTT.....TCTGAAGAGACCCCT 1256
438 rSerGlyAspAspAlaCysSerPheSerGlyPheArgProAlaThrLeu 455
1257 CTCCTTGAGAGAAATGGGTTGATCCAACTTCAAAACCTCCACTCTGA 1306
455 hrValThrAsnPhenPheGlyGlnGluAlaGluArgLeuSerAspGlu 471
1307 GCGTTAGCAGCTTTTTCACAGCAGAAAGAGATGCTTACGGATGAAGAC 1356
472 LeuPheLysPheLeuAlaAspMetArgArgProSerSerLeuLeuArg 488
1357 TTTATTCAGTTTAACTGACTACAAAAGATCATCATCTTACAGAGACG 1406
488 GLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAlaPro 505
1407 AGTCAACTCAATTCAGGCTTGCTAGACAGTGAATTTCTACAGCTCCAG 1456
505 LysAsnProHisPheCysLeuSerProGlnLeuLeuHisIleLysPro 521
1457 AGATCATCAATTCGTGTGACTGCTGAATTCGCTCCCGTGAAGACCTTT 1506
522 ProAspProArgGlyArgProThrLysGluIleLeuGlnPheProAla 538
1507 CTGGAAGAACGGACAGCCCGCCACAAAGATTTTGGAAATTTCCACACG 1556
538 gGluValThrAlaProHisThrSerTyrArgAsnLeuLeuTyrValTyr 555
1557 AGAAGTATGTCCTCACACTGTGACAGAAACCTTCTATGCTTACC 1606
555 roHisSerLeuAsrPheSerSerArgGlnGlySerValArgAsnLeuAla 571
1607 CACGAGGCTGAACTTTGTAAACAAACTACAGACCGCGAATTTACA 1656
572 ValArgValGlnTyrMetThrGlyGluAspProSerGlnAlaLeuPro 588
1657 ATAAAGATCCAGTTTATGTGTGAGAAAGATCGTACATGCGATGCCGG 1706
588 lIlePheGlyLysSerSerCysSerGlnPheThrArgGluAlaPheThr 605
1707 CACTTTGGAAATCCAGCGGCTGATTTCTGAGAAAGTCTACACAG 1756
605 roValValTyrHisAsnLysSerProGlnPheTyrGlnGluPheLys 621
1757 CTGTTACATACCAATTAAGTCTGACTTGTATTAAGAAGTGAATAATT 1806
622 HisLeuProAlaCysValThrGluAsnHisIleLeuPheThrPhe 638
1807 AAGTCCCGCTAAGCTACAGTAATACACACCTCTGTCCACTTC 1856
638 rHisValSerCysGlnProArgProGlyThrAlaLeuGlnTyrProVal 655
1857 CCATATCAGCTGTACAGCAAGCAAGAGGCTCCGTGAAACTCTCTG 1906
655 lPheThrTyrPheProLeuLeuGlnHisGlyArgLeuArgThrGlyPro 671
1907 GATATTCATGCTGCCAATTCCTTAATGAAGCTTCAACCTGGATCC 1956
672 PheCysLeuProValSerValAspGlnProProSerTyrSerVal 688
1957 TACTGTCTCCAGTGTGCTTGGAAATTTGCCACCACTACTCCATGCA 2006
688 uHisProAsp.....ValAlaLeuProGlyMetArgTyrProValAsp 702
2007 TCTGTGCAAGAAAGTCCCATTTACAGAAATCTCCCATTAAGTGGCTGA 2056
702 lHisLysGlyValPheSerValGlnLeuThrAlaValSerSerValHis 718

```

```

2057 GACATPAGGAGATTTATATTAATGAATGCAAGCTTTCTGTGTACAC 2106
719 ProGlnAspProTyrLeuAspLysPhePheThrLeuValHisValLeu 735
2107 ACCCAGCAACACCACCTGGAGAGATTTCAOCCCTCTCCACTCCCTGGA 2156
735 u...GluGlyAlaPheProPheArgLeuLysAspThrValLeuSerGlu 751
2157 GAGCCAGGAGACCTTCCCATCCGCGTGTGATCAGAAATCAGACGGAGA 2206
751 LysAsnValGlnGlnLeuArgAlaSerLeuAlaLeuArgLeuAla 767
2207 TGGCCTGAGCATGAGCTGAAGCTCAGCATATCTGCTTAACCTCTCC 2256
768 SerProGlnProLeuValAlaPheSerHisValLeuAspLysLeuVal 784
2257 GCGCTGAGAGCGCTGTGCTCTCTGACACCTGGTGGTGGCAAGCTCTT 2306
784 lArgLeuValIleArgProProIleIleSerGlyGlnIleValAsnLeu 801
2307 CCAGCTGTCCGTGACAGCCCATGCTCATGCTGGCCGACAGCAACTTCT 2356
801 lArgGlyAlaPheGlnAlaMetAlaHisValValSerLeuValHisArg 817
2357 CCAGTGTGCTTGAGTCCGAGTCCGTGTGCGCATGCGCAAGTCTGACAC 2406
818 SerLeuGluAlaAlaGlnAspAlaArgGlyHisCysProGlnLeuAla 834
2407 AGCAAGACACTGACAGAGACAGCATGGAGAACCTCCCTGCTGCTTC 2456
834 aTyrValHisTyrAlaPheArgLeuProGlyThrGlnProSerLeuPro 850
2457 CTACGTGCACTACGTCTTCCGCTGCCAGAGTGCAGAAAGGATGTGCCCA 2506
851 ..AspGlyAlaProProValThrValGlnAlaAla.....ThrLeu 863
2507 ACTCAGGCGCTCCACTGCTCCTAGACCTCGGAGCTTACCACACGTAT 2556
864 AlaArgGlySerGlyArgProAlaSerLeuTyrLeuAlaArgSerLys 880
2557 GCGCCGACATCAGCTGCTGTGTGATTCAAAGCTGCTGACAGCGCGGCT 2606
880 rIleSerSerSerAsnProAspLeuAlaValAlaProGlySerValAsp 897
2607 GATGAGCAGCATTAACCACTCGCGGGACACACTCCGACAGACAGCG 2656
897 spGluValSerArgIleLeuAlaSerLysLeu..... 907
2657 AGGAAGTGAAGAACATCATGCTTCAAGATCGCGATCGCACTGCAGC 2706
907 ..... 907
2707 CGAATGCTTACTATTGCTCTGCGAGTAGTGATGCTCCAACTGCACG 2756
908 ..... 908
2757 AGCCCAAGAGCCAGCCAGCAAAAAGCATTTCCATGAGAGAGCTTGCC 2806
915 lIntProValValSerSerSerAlaValArgGluAlaIleLeuGlnHisAla 931
2807 AGATGTGTGTCAGACCGGA..... 2826
932 TrpPhePhePheGlnLeuMetValLysSerMetAlaLeuHisLeuLe 948
2827 ..... 2827
948 uGlyGlnAlaArgLeuAspThrProArgLysLeuArgPheProGlyArgPhe 965
2859 CATGGAACAACGCGACATTTTCGAGAGACATCGTTTCTGACCGCTTCA 2908
965 euAspAspIleThrAlaLeuValGlySerValGlyLeuGlnVal..... 979

```

2909 TGGATGACATACACTATTGTTAATGTGGTCACCTCGAATAATGCAGCC 2958  
980 ... ILeThrArgValHisLysAspValGluLeuAlaGluHisLeuAsnAl 995  
2959 CTTTAGTAAACACACAGAAAGAAATGMAACGGCGAAAGATGAAACAT 3008  
995 aSerLeuAlaPhePheLeuSerAspLeuSerLeuValAspArgGlyP 1012  
3009 CAGCTGCTCTTCTTGTGTGACCTTCTCTCCCTCATGGATCGGGGCT 3058  
1012 heValPheSerLeuValArgAlaHisTyrLysGlnValAlaThrArgLeu 1028  
3059 TTGGTTAACTCATCAGACATTATGAGCCAGCTGTGACGCCAACCTC 3108  
1029 GlnSerSerProAsnProAlaAlaLeuLeuThrLeuArgMetGluPheTh 1045  
3109 AGTAACTTCA.....ACGCTATTTCCTGAGGCTAGAGTTCCCT 3149  
1045 rArgIleLeuGlySerHisGluHisTyrValThrLeuAsnLeuProGlySc 1062  
3150 GAGATCTCTGTAGCCATGAGCATTTACCTCAATCTGAACCTTTTCTTA 3199  
1062 ySProLeu...SerProAlaSerProSerProSerValSerSerThr 1077  
3200 TGAATGCTGATCTGCTCCACATCTCTCTCTCCATATCT..... 3243  
1078 ThrSerGlnSerSerThrPheSerSerGlnAlaProAspProLysValTh 1094  
3244 ...TCCAGAACCTCAAGCTCTGCTCCAGCTTCACAGAACCAAGATCGC 3290  
1094 rSerMetPheGluLeuSerGlyProPheArgGlnGlnHisPheLeuAlaG 1111  
3291 CAGATGTCTCATCTGACTTCGCGATCCGACCGCACACAGCTTCTCCACGC 3340  
1111 LysLeuLeuLeuThrGluLeuAlaLeuAlaLeuGluProGluAlaGluGly 1127  
3341 GGCTCTCTTCACAGAACTGGCTGCTGCTCGATGCGCAAGGGAAGGA 3390  
1128 AlaPheLeuLeuHisLysLysAlaIleSerAlaValHisSerLeuLeuCy 1144  
3391 ATCAGAAAGTACAAAGAAAGCTGTGCAATTCACAGCTGTGCAAG 3440  
1144 sGlyHisAspThrAspProArgTyrAlaGluAlaThrValLysAlaArgV 1161  
3441 TTCACGACCTGAGCCAGCTGTGTCAAAACACAGAGTGAAGGTCAAAA 3490  
1161 alaAlaGluLeuTyrLeuProLeuLeuSerIleAlaArgAspThrLeuPro 1177  
3491 TCGCGCGCTTACCTACCTTACCTTACCTTACCTTACCTTACCTTACCA 3540  
1178 ArgLeuHisAspPheAlaGluGlyProGlyGlnArgSerArgLeuAlaSe 1194  
3541 CAGCTCTGTGCTTACAGTTGAGATACGCGAGATACCGCACACAGCTGG 3590  
1194 rMetLeuAspSerAspThrGluGluGluGlyAspIleAlaGlyThrIleA 1211  
3591 C.....TCGGATGAAGAACCAAGAAAGA.....GCCGGTGCACATTA 3625  
1211 snProSerValAlaMetAlaIleAlaGlyGlyProLeuAlaProGlySer 1227  
3626 ACCGAAATGTGGCTCTGCGCATACAGGGAATATTTCAATTTGAACACA 3675  
1228 ArgAlaSerIleSerGlnGlyProProThrAlaSerArgAlaGlyCysAl 1244  
3676 AGTGAATAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3719  
1244 aLeuSerAlaGluSerSerArgThrLeuLeuAlaCysValLeuTyrVal 1261  
3720 GCTGAACGCGACACTACTGCAACCTCATGATCTGCTCTCTGATCA 3769  
1261 euLysAsnThrGluProAlaLeuGlnGlnArgThrAlaThrAspLeuThr 1277  
3770 TGAATAATGCTGATCAGACCTCTATTAGGAATGGAATTTGCTGACCTGCA 3819

1278 LeuProGlnLeuGlyArgLeuLeuAspLeuLeuTyrLeuGlyLeuAlaAl 1294  
3820 TCAACCCACTCATCAGAGATTTTAGATCTACTTTCATCTGTGTATG 3869  
1294 abPheGluTyrLysGlyLysLysAlaPheGluArgIleAsnSerLeuThrP 1311  
3870 TTTTGGATTAAGGGGAAACACAGATCTGACAAAGTCAAGTACCAAGTCC 3919  
1311 heLysLysSerLeuAspMetLysAlaArgLeuGluGluAlaIleLeuGly 1327  
3920 TGCAGAGTCAAGGGATGTCAAGGCCGCTGGAAGAGCTTTGCTGCTG 3969  
1328 ThrIleGlyAlaArgGlnGluMetValArgArg.....SerArgI 1341  
3970 GGGGAAGGGGCGAGAGGAGATGATGCGCCGCGGCTCCAGGGAACA 4019  
1341 uArgSerProPheGlyAsnProGluAsnValArgThrArgLysSerValT 1358  
4020 CCGATTTCCA...GGCCTAAATGAATAATTTGAGATGGAAGAAAGACAGA 4066  
1358 hrHisTyrLysGlnThrSerAspArgValAspLysThrLysAspGluMet 1374  
4067 CACATTGGCGGCGACAGCTAATGAAAGCTAGATAAACAAAGCCGAGTTA 4116  
1375 GluHisGluAlaLeuValGluGlyAsnLeuAlaThrGluAlaSerLeuVa 1391  
4117 GATCAAGAAAGCTTGATCATGAGTCAATCTGGCTACAGAAACATTTAAT 4166  
1391 ValLeuAspThrLeuGluIleIleValGlnThrValMetLeuSerGluA 1408  
4167 CATCTGGATATGACAGAAACATTATTCAGGCGAGCTCGCTGTGACT 4216  
1408 IaArgGluSerValLeuGluAlaValaLeuLysValaValLeuTyrSerLeu 1424  
4217 GTAAGACAGCTGCTGGAGGAGTGTCTGAGGCTCTGTAATCTCTG 4266  
1425 GlySerAlaGlnSerAlaLeuPheLeuGlnHisGlyLeuAlaThrGlnAr 1441  
4267 AACTGTGACAGAGTACACCTCACTGACTCACTCTTTCGACACCTCGC 4316  
1441 gAlaLeuValSerLysPheProGluLeuLeuPheGluGluAspThrGluL 1458  
4317 TGTCTCATGCGCAAGTTTGGAGACTTACTCTTCAAGAGGAGGTGGAAAC 4366  
1458 euCysAlaAspLeuGlySerLeuArgLeuLeuArgHisCysGlySerArgIle 1474  
4367 AGTGTTTGACCTATGTACACAAAGTCTGACACACTGACAGCAGCATG 4416  
1475 SerThrIleArgThrHisAlaSerAlaSerLeuTyrLeuLeuMetArgG 1491  
4417 GATGTACCCGAGGACCAAGCTGTCACACCTTACCTTCATCATGAGGTT 4466  
1491 nasnPheGluIleGlyHisAsnPheAlaArgValaLysMetGlnValThrM 1508  
4467 CAGTTTGGAGCCACAGTAATTTTGGCAAGATAAGATGCAAGTAACCA 4516  
1508 eSerLeuSerSerLeuValGlyThrThrGlnAsnPheSerGluGlnHis 1524  
4517 TGTCCCTGCGCATCTTGTGGGGAAGACACACAGACTTAAATGAAGACAC 4566  
1525 LeuArgArgSerLeuLysThrIleLeuThrTyrAlaGluGluAspMetG 1541  
4567 CTGAGAAAGATCTTGAGACAAATTTTGGCTATTCAGAAAGACACAGC 4616  
1541 yLeuArgAspSerThrPheAlaGluGlnValaGlnAspLeuMetPheAsnL 1558  
4617 CATGAGATGATCTCTTCCACCCAGCTGAGAACTTCTGTATATC 4666  
1558 euHisMetIleLeuThrAspThrValaLysMetLysGlnGluGluAsp 1574  
4667 TGAATAATGCTTATATGACACAGTGAATAATGAGGAATTTTCAGGAAGAT 4716



```

1575 ProGluMetLeuIleAspLeuMetYrArgIleAlaArgGlyTyrGlnG 1591
|||||.....|
4717 CCTGAGATGCTTATGAGATCTCATGTACAGAAATGCCAAGAGTACCAAGC 4766
|||||.....|
1591 ySerProAspLeuAvrLeuThrTrpLeuGlnAsnMetAlaGlyYrHisA 1608
|||||.....|
4767 ATCTCTGATCTGCGCTGACCTGCTCCAGAACTGGCAGAGAAACACA 4816
|||||.....|
1608 IaGluLeuGlyAsnHisAlaGluAlaAlaGlnCysMetValHisAlaAla 1624
|||||.....|
4817 CCAAGAGAGAGAGTGTACACGGAGGCTGCCATGTGCTGTGCACGCCCT 4866
|||||.....|
1625 AlaLeuValAlaGluTyrLeuAlaLeuLeuGlnAspGlnArGHisLeuP 1641
|||||.....|
4867 GCGTTAGTGGCGCTGAGTATCTGAGCACTGTGAGAGCACACACTTACTGCC 4916
|||||.....|
1641 oValGlyCysValSerPheGlnAsnIleSerSerAsnValLeuGlnGln 1658
|||||.....|
4917 CGTGGGCAGTGCTGACCTCCAGAAATATTCTTCCAAATGCTGGAGAGAGT 4966
|||||.....|
1658 eValIleSerAspAspIleLeuSerProAspGlnGlnGlyPheCysSer 1674
|||||.....|
4967 CTGTGCTCTGAGGACACCCCTGTCTACTGACGAGAGATGGGCTGTGCCCA 5016
|||||.....|
1675 GlyYrHisPheThrGluLeuGlyLeuValGlyLeuLeuGlnGlnAlaAl 1691
|||||.....|
5017 GCCCAGTACTTACCCGAGAGATGGCGCTGTAGCCCTCTGGAGACAGCCGC 5066
|||||.....|
1691 aGlyYrPheThrMetGlyGlyLeuTyrGluAlaValAsnGluValTyrL 1708
|||||.....|
5067 GGAGCTCTGACGACGGAGGCTTATATAGACAGTTAATGAGGTCTTACA 5116
|||||.....|
1708 yAsnLeuIleProIleLeuGlnAlaHisArgAspTyrLysLysLeuAla 1724
|||||.....|
5117 AGCTGTGATCCCATCTCTAGAGCGCATCGAGAAATTCGGAGAGGTGACA 5166
|||||.....|
1725 AlaValHisGlyLysLeuGlnGlnAlaPheThrYrIleMetHisGlnSe 1741
|||||.....|
5167 CTCACTGCACAGCAAGCTCGACAGAGACCTTCGACAGCATCTTAACAAGCA 5216
|||||.....|
1741 rSerGlyTrpGluTyrValPheGlyThrTyrPheArgValGlyPheYrG 1758
|||||.....|
5217 TCAT.....AAGAGAAATGTTTGAACCTTCTCCGAGTGTGTTTCTTGG 5260
|||||.....|
1758 lYAlaHisPheGlyAspLeuAspGlnGlnGluPheValTyrLysGluPro 1774
|||||.....|
5261 GATCCAAATTTGGGGATTTGGATGACAGAGGTTTGTCTAACAAAGAGCCT 5310
|||||.....|
1775 SerIleThrLysLeuAlaGluIleSerHisArgLeuGlnGluPheYrTh 1791
|||||.....|
5311 GCAATTACCAAGCTTCTGAGATCTCAATAGACTAGAGGATTTTATGG 5360
|||||.....|
1791 rGluArPheGlyAspValValGluIleIleLysAspSerYrProV 1808
|||||.....|
5361 TCATATTGTTGTGTCAGAAATTTGGGAATGATTAAAGACTCCACTCTCG 5410
|||||.....|
1808 aLAspLysSerLysLeuAspSerGlnLysAlaTyrIleGlnIleThrTyr 1824
|||||.....|
5411 TGGACAAAGAACAGTGGATCTTAACAAGGCTCATACAGATCACTTTT 5460
|||||.....|
1825 ValGluProTyrPheAspThrTyrGluLeuLysAspArgValThrTyrPh 1841
|||||.....|
5461 GTGAGAGCCCTTCTTGAATGATGATGAAAGACAGGAGGTCACATACTT 5510
|||||.....|
1841 eAspArgAsnTyrCylLeuArgThrPheLeuPheCysThrProPheThrP 1858
|||||.....|
5511 TGAAGAGAAATTTCAACCTCCGGAGTTCATGTACACCAACCCCTTCAACC 5560
|||||.....|
1858 roAspGlyArgAlaHisGlyGluLeuProGlnGlnHisLysArgLysThr 1874
|||||.....|
5561 TGGAGGGGCGGCGCTCGGGAGAGCTGATGAGCAGTACAGAAAGAACACA 5610
|||||.....|
1875 LeuLeuSerThrAspHisAlaPheProTyrIleLysThrArgIleArgVa 1891

```

```

5611 GTCTTGACCACTATGACAGCGCTTCCCTACATCAAGACCGATCAGCGT 5660
|||||.....|
1891 lCysHisArgGlnGluThrValLeuThrProValGluValAlaIleGluA 1908
|||||.....|
5661 CATCAACAAGAGAGAGTGTGTTTGGACCGGATTTGAAGTTGCCATTGAAG 5710
|||||.....|
1908 sPheGlnLysLysThrArgGluLeuAlaPheAlaThrGlnGlnAspPro 1924
|||||.....|
5711 ACATGAAGAAGAAGACCCCTGACATTAGCACTTCCATTACCGAGAGCCG 5760
|||||.....|
1925 ProAspAlaLysMetLeuGlnMetValLeuGlnGlySerValGlyProH 1941
|||||.....|
5761 CCGATGCAAGAGATGCTTCAGATGTGCTGCAAGGCTCTGTGGAGCTAC 5810
|||||.....|
1941 rValAsnGlnGlyProLeuGlnValAlaGlnValPheLeuAlaGluIleP 1958
|||||.....|
5811 TGTAATCAAGAGACCACTGAGATAGCCCAAGTGTGTTGGCTGAATATTC 5860
|||||.....|
1958 roGluAspProLysLeuPheArgHisHisAsnLysLeuArgLeuCysPhe 1974
|||||.....|
5861 CTGCTGATCCAAAACCTATACGACATCACAAACAGTTGAGGTTATCTTT 5910
|||||.....|
1975 LysAspPheCysLysLysCysGluAspAlaLeuArgLysAsnLysAlaLe 1991
|||||.....|
5911 AAGGAATTCATCATGAGATGTGTGAAGCTGTAGAGAAAACAGGCTCT 5960
|||||.....|
1991 uIleGlyProAspGlnLysGluTyrHisArgGluLeuGlnArgAsnTyrC 2008
|||||.....|
5961 CATCAGCGGACAGCAGAGGAGATATCAGCAGAACTCAAAAAGAACTATA 6010
|||||.....|
2008 yAsArgLeuArgGluAlaLeuGlnProLeuLeuThrGlnArgLeuProGln 2024
|||||.....|
6011 ACAAGCTAAAAGAAACCTCAGGCCATGATCGAGGGGAAAATTCAGAA 6060
|||||.....|
2025 LeuMetAlaPro.....ThrProGlyLeuAlaArgAsnSerLeuAs 2038
|||||.....|
6061 CTGACACAGCCAAATATTCAAGATTGAGAGTCAAAAGAGGACTCTTCCA 6110
|||||.....|
2038 nArgAlaSerPheArgLysAlaAsp 2046
|||||.....|
6111 CAGATCTAGTTTCAGGAAATGTGAA 6135
|||||.....|
seq_name: /sids1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:ABLI17736
seq_documentation_block:
ID ABLI17736 standard; DNA; 9389 BP.
XX
XX ABLI17736;
AC
XX
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster genomic polynucleotide seq ID NO 4681.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW
XX pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW.
PI
XX
XX WPI; 2001-656860/75.
DR

```







3271 GCGGTGGCTTAAACCGTCTGTTATGCTGATCCGACCTCCGACTG 3222  
1543 ArgAspSerThrPheAlaGluGlnValGlnAspLeuMetPheAsnLeuH1 1559  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
3221 CAGGATACCTCTTCTCTGCAACAGTACAAAGATTCTTTTAACTGCA 3172  
1559 SMetIleLeuThrAspThrValLysMetLysGluHISGlnGluAspProG 1576  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
3171 TATGATCCCTGTCAGATCTGTTAAATGAAGGATGTACAGGAAGACCCAG 3122  
1576 LuMetLeuIleAspLeuMetTyrArgIleAlaArgIleTyrGlnGlySer 1592  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
3121 AATGTGCTGCTGACCTCATGAAATCGTATGCGCAAGGATACCAAAATAC 3072  
1593 ProAspLeuArgLeuThrTyrPleuGlnAsnMetAlaGlyLysHisAlaG1 1609  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
3071 CCTGATCTGAGCTGACTGTTGGTGAATAATGCTAAAAAACCCGCGCA 3022  
1609 uLeuGlyAsnHisAlaGluAlaAlaGlnCysMetValHisAlaAlaAlaL 1626  
|:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
3021 GCGACGAAATCACACGGAAGACGCCATGTGTATGTACATGCTGCTTCT 2972  
1626 euValAlaGluTyrLeuAlaLeuLeuGluAspGlnArgHisLeuProVal 1642  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
2971 TAGTTCTGAAATCTTAGCATGTGTGAGTCACAAAACATTTGCCCTGT 2922  
1643 GlyCysValSerPheGlnAsnIleSerAsnValLeuGluGluSerAl 1659  
|:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
2921 GGAGCTGTAAGTTTTCACGAAATTTCTCCCAACACACTAATGAGCGCGC 2872  
1659 aIleSerAspIleLeuSerProAspGluGluGlyPheCysSerGlyL 1676  
|:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
2871 CGATGCGATGATGCTGTAAGTCCCGCAAGATGTATCTGCTTAGGAA 2822  
1676 yHisPheThrGluLeuGlyLeuValGlyLeuLeuGluGlnAlaAlaGly 1692  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
2821 ATCATTTCACTGAACCTGGGTGAGGCTTGCCTGGAAGAGCCCTCATT 2772  
1693 TyrPheThrMetGlyLeuTyrGluAlaValAsnGluValTyrLysAs 1709  
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
2771 TCTTTCAGAGTTGCTGCATGATGAAGCAATGACGAAGTGTACAAAT 2722  
1709 nLeuIleProIleLeuGluAlaHisArgAspTyrLysLysLeuAlaAlaV 1726  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
2721 TCTAAATACCATATGCGAGGCTAACAGAGATTTCAAAAGCTAAGCAAG 2672  
1726 aHisGlyLysLeuGlnGluAlaPheThr..... 1735  
2671 TTCATGGCAAAATGCAAGAGGCTTAATCGAATATCCCACTACAGTA 2622  
1736 .....LysIle.MetHisGlnSerSerg 1743  
2621 ACAATATGTGTAAATTTTACCAACGAAACATATATATTTATAAAC 2572  
1743 LyrProGluArgValPheGlyThrTyrPheArgValGlyPheTyrGlyAla 1759  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
2571 AGGCTAAGAGAGTTTGGAACTACTTTCGTGCTTCTTAGTGGCGGA 2522  
1760 HisPheGlyAspLeuAspGluGlnGluPheValTyrLysGluProSerL1 1776  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
2521 AAATTTGGGACTT3GATCAGCAGGAATTCATTATTAAGAGCAACATT 2472  
1776 eThrLysLeuAlaGluIleSerHisArgLeuGlu..... 1787  
2471 GACGAAGTTGCGCAAAATTTTATGTCGCTCAGGTATATATTTGCAATT 2422  
1788 .....Glu.PheTyr 1790  
2421 GGAATAAATAGAACTATCATTTTCTTCAACATACGTTAGACTTTTA 2372  
1790 rThrGluArgPheGlyAspAspValAlaGluIleIleLysAspSerTyrP 1807  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
2371 CACTGAACGATTCGACGCGACTCTGTCATATCATTAAGATTCACATA 2322

1807 roValAspLysSerLysLysAspSerGlnLysAlaTyrIleGlnIleThr 1823  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
2321 CCGTGAATATTAAATAGCTTGGATCCCGATTAAGGCTACATTCAAATTACT 2272  
1824 TyrValGluProTyrPheAspThrTyrGluLeuLysAspArgValThrTy 1840  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
2271 TATGTTGAACCTTACTTTGAAACATATGAAATGCGTCATGCTGAGACATA 2222  
1840 rPheAspArgAsnTyrGlyLeu..... 1847  
2221 CTTTACGCGGAATTCATATATAGATGATATGATTAACAGATAATT 2172  
1848 .....ArgThr 1849  
2171 AAATGAAATTTAAATATATATTATTAACATTTTGGTTATGCAAGAC 2122  
1850 PheLeuPheCys..ThrProPheThrProAspGlyArgAlaHisGlyGlu 1866  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
2121 TTTTATATATGCGACGCTTTTACTAAAAAATGAAGGACACATGCGCAAC 2072  
1866 euProGluGlnHisLysArgLysThrLeuLeuSerThrAspHisAlaPhe 1882  
|||:::|||||:::|||||:::|||||:::|||||:::|||||  
2071 TAAATGACACAGTGTAAACGAAACGATTTTGACCGCGGCAATCACTTT 2022  
1883 ProTyrIleLysThrArgIleArgValCysHisArgGluThrValLe 1899  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
2021 CCATCGTAAAAACCGCATATGTTATGTAATAGCCGACAGCAAAATAGTCT 1972  
1899 uThrProValGluValAlaIleGlu..... 1907  
1971 GCAACCGATTGAGTGGCAATTTGAAGTATGTTGTTAAAAATCCCAA 1922  
1908 .....AspMetGlnLysLysThrArg 1914  
1921 AGTATATAAATTAACGTTATTCCTAGATATTTCAAAAAAACATGTG 1872  
1915 GluLeuAlaPheAlaThrGluGlnAspProProAspAlaLysMetLeuG1 1931  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1871 GAATGGCAGCTGCCCAATCAACAGAGCCAGACAGCCCAAAATATGCA 1822  
1931 nMetValLeuGlnGlnLysSerValGlyProThrValAsnGlnLysProLeuG 1948  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1821 AATGCTACTGCAAGGATGTATTTGACCACTGTTTACCAGAGCCGATGG 1772  
1948 luValAlaGlnValPheLeuAlaGluIleProGluAspProLysLeu... 1963  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1771 AAATGCGAGATGTCTCTTCCAAATTTATCCAGAGCAACATGTAACG 1722  
1964 PheArgHisHisAsnLysLeuArgLeuGlyPheLysAspPheCysLysLy 1980  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1721 ACAAACACCAAAACAAACTTCGGTGTGCTTGCAGGATTTTCAAGCGC 1672  
1980 sCysGluAspAlaLeuArgLysAsnLysAlaLeuIleGlyProAspGlnL 1997  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1671 TTGTCTGATGCTTTGAAAGAAAGAAATCGCAATCTATATCTTCAAGTCAA 1622  
1997 ySGluTyrHisArgGluLeuGluArgAsnTyrCysArgLeuArgGluAla 2013  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
1621 AAGATTACCAACGAGAACTGAGAGCTAACAGCATCGGTTCAATTGAACGA 1572  
2014 LeuGlnProLeuLeuThr 2019  
|||:::|||||:::|||||:::|||||:::|||||:::|||||  
1571 TTGACTCCCTTATATACT 1554  
seq\_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.AAH41953  
seq\_documentation\_block:  
ID AAH41953 standard; DNA: 4200 BP.  
XX AAH41953;  
AC  
XX  
DT 30-AUG-2001 (first entry)



1109 uAlaGlyLeuLeuLeuThrGluLeuAlaLeuAlaLeuGluProGluAlaG 1126  
|||||.....|||||..... 1129 |||||.....  
947 GCGAGGACTGTGTATACAGAGCTGGCTGCATTTAGACCCCTGATGCTG 996  
1126 LuGlyAlaPheLeuLeuHisLysAlaIleSerAlaValHisSerLeu 1142  
|||||..... 1129 |||||.....  
997 AAGAGCTGTTGGATTGCATTAAGAAAGTCATCAATATGTGTACAACTTA 1046  
1143 LeuGlySerLysHisAspThrAspProArgTyrAlaGluAlaThrValLysAl 1159  
|||..... 1129 |||||.....  
1047 CTCCTCCAGTACGAGCTCAGACCCGGGTACTCTGACCCCTCAGATTAAGGC 1096  
1159 AArgValAlaGluLeuTyrLeuProLeuLeuSerIleAlaArgAspThrL 1176  
|||||..... 1129 |||||.....  
1097 TCGAGTGGCCATGTTGTATCTACCTCGATTTGGATTATCATGAAACCTG 1146  
1176 euProArgLeuHisAspPheAlaGluGlyProGlyGlnArgSerArgLeu 1192  
:|||||..... 1129 :|||||.....  
1147 TACCTCAGCTGTATGATTTTACAGAAACTCCAAATCAACGAGAGAACCA 1196  
1193 AlaSerMetLeuAspSerAspThrGluGlyGluGlyAspIleAlaGlyTh 1209  
:|||||..... 1129 :|||||.....  
1197 ATTGTGTATGCCACTGATGATTAAGAAAGTGAGAGC.....GGAAGTAT 1240  
1209 rIleAsnProSerValAlaMetAlaIleAlaGlyGlyProLeu..... 1223  
:|||||..... 1129 :|||||.....  
1241 GATTAAGCCAGACCGTGGCATGCGCATCGCAGGACATCGGTCCCTCAAC 1290  
1224 .....AlaProGlySerArgAlaSerIleSerGlnGlyProProThrAla 1238  
|||||..... 1129 |||||.....  
1291 TAAACAGCCCTGGCAGATTCTCTCTCAGCTCA.....ACGAGT 1338  
1239 SerArgAlaGlyCysAlaLeuSerAlaGluSerSerArgThrLeuLeuAl 1255  
:|||||..... 1129 :|||||.....  
1339 GCGAGGACACACTACTCTTTCAGCAGAAATCAACAGTCGAAACCTTTGAT 1378  
1255 aCysValLeuTyrValLeuLysAsnThrGluProAlaLeuLeuGlnArgT 1272  
|||||..... 1129 |||||.....  
1379 CTGTCTACTTGGGTTCTCAAAAATGCAGATGAACAGTTCTACAGAAAGT 1428  
1272 rPalaThrAspLeuThrLeuProGlnLeuGlyArgLeuLeuLeuLeu 1288  
|||||..... 1129 |||||.....  
1429 GGTTTACAGATCTCTCAGCTTTCGAGCTAAACCGGCTATTAGATCTGCTT 1478  
1289 TyrLeuCysLeuAlaAlaPheGluTyrLysGlyLysAlaPheGluAr 1305  
|||||..... 1129 |||||.....  
1479 TATCTCTGTGTCTTCTGCTTGGATATAAGGAAAGGTTTGAACG 1528  
1305 gIleAsnSerLeuThrPheLysLysSerLeuAspMetLysAlaArgLeuG 1322  
:|||||..... 1129 :|||||.....  
1529 AATGAATAGCTTGACCTTAAAGAAATCAAAAGACATGAGACAAAGCTTG 1578  
1322 LuGluAlaIleLeuGlyThrIleGlyAlaArgGlnGluMetValArgArg 1338  
|||||..... 1129 |||||.....  
1579 AAGAAGCTATTCTTGGAGCATAGGTGCCAGCAAGAAATGCTACGGCGCA 1628  
1339 SerArg.....GluArgSerPro.....PheGlyAsnPr 1348  
|||||..... 1129 |||||.....  
1629 AGCCGAGGACAGCTCGAGAGAAAGCCCATCTGGAAGTGCCTTGGAAAGTCA 1678  
1348 oGluAsnValArgTyrPArgLysSerValThHisTyrLysGlnThrSerA 1365  
|||||..... 1129 |||||.....  
1679 AGAAAATTGTAGGTGAGAGAAAGATATACACGCGGTCAAAAACACAG 1728  
1365 sPArgValAspLysThrLysAspGluMetGluHisGluAlaLeuValGlu 1381  
:|||||..... 1129 :|||||.....  
1729 AGAAGCTTGACAAATCAAGACGAGATTTGAACACGAAACACTGATTGAT 1778  
1382 gIyAsnLeuAlaThrGluAlaSerLeuValValLeuAspThrLeuGluI 1398  
|||||..... 1129 |||||.....  
1779 GGAAACCTGGCTACAGAAAGCAAACTAATCATTTAGATACATTAGAGAT 1828

1398 eIleValGlnThrValMetLeuSerGluAlaArgGluSerValLeuGlyA 1415  
|:|:|..... 1399 :|:|:|.....  
1829 TGTGTTCAGACCGTTTCTGTAAACCGAAATCCAAAGACATTTCTTGCTG 1878  
1415 lValLeuLysValValLeuTyrSerLeuGlySerAlaGlnSerAlaLeu 1431  
:|||||..... 1399 :|||||.....  
1879 GAGGTCTAAAGTGTCTACTACACAGCATGGCTGTAAACCAAGTCCAGTT 1928  
1432 PheLeuGlnHisGlyLeuAlaThrGlnArgAlaLeuValSerLysPhePr 1448  
:|||||..... 1399 :|||||.....  
1929 TATCTAACAACACTGTTTCTCTACACAGAGAGCCCTGTTCAGAAAGTTTCC 1978  
1448 oGluLeuLeuPheGluGluAspThrGluLeuGlyAlaAspLeuCysLeuA 1465  
|||||..... 1399 |||||.....  
1979 TGAACCTTATTGTGAAGAGACAGACAGAGTGTCTGATTTATATCTTCA 2028  
1465 rGluLeuArgHisCysGlySerArgIleSerThrIleArgThrHisAla 1481  
|||||..... 1399 |||||.....  
2029 GCGTTCTCGACACTGTATACACTAGCATCGTACAAATACGGTCACACCCC 2078  
1482 SerAlaSerLeuTyrLeuLeuMetArgGlnAsnPheGluIleGlyHisAs 1498  
:|||||..... 1399 :|||||.....  
2079 AGTGCCTCCCTTACTACTAATGAGGCAAACTTTGAGATGGGAAATGA 2128  
1498 nPheAlaArgValLysMetGlnValThrMetSerLeuSerSerLeuValG 1515  
|||||..... 1399 |||||.....  
2129 CTTTGCAGGGTTAAATACAGGTACCAATGTCACTATCTCTCTTGCTG 2178  
1515 lYThrThrGlnAsnPheSerGluGlnHisLeuArgArgSerLeuLysThr 1531  
|||||..... 1399 |||||.....  
2179 GCACTACTCAGAAATTTTAAAGAAGATCTTAAGACGTTCTCTAAAGACT 2228  
1532 lIleLeuThrTyrAlaGluGluAspMetGlyLeuArgAspSerThrPheAl 1548  
:|||||..... 1399 :|||||.....  
2229 ATATTGACATATGCTGACAGAGATTCGAATTGAGGAAACAACATTTCC 2278  
1548 aGluGlnValGlnAspLeuMetPheAsnLeuHisMetIleLeuThrAspT 1565  
:|||||..... 1399 :|||||.....  
2279 TGAATCAGTCCAGAGATCTGTTTCAATCTCCATATGATTTCTTTCGATA 2328  
1565 hValLysMetLysGlnHisGlnGluAspProGluMetLeuIleAspLeu 1581  
CTGTGAAGAAATGAAGGAACACAGAGAGATCCGTAATGTGATGATGATCA 2378  
1582 MetTyrArgIleAlaArgGlyTyrGlnGlySerProAspLeuArgLeuTh 1598  
|||||..... 1399 |||||.....  
2379 ATGTACGAATTCGCAAGGGTTACCAAGCTCTCCAGATCTGGATTTGAC 2428  
1598 rTyrPLeuGlnAsnMetAlaGlyLysHisAlaGluLeuGlyAsnHisAlaG 1615  
|||||..... 1399 |||||.....  
2429 CTGGTTCCAGAACATGCGACAGCAAGCACTCAGAACGAAATCATGCTG 2478  
1615 LuAlaAlaGlnCysMetValHisAlaAlaAlaLeuValAlaGluTyrLeu 1631  
|||||..... 1399 |||||.....  
2479 AAGCTGCACAGTGTCTAGTCCACTCAGACACTGTGTGCTGTAATTTTG 2528  
1632 AlaLeuLeuLysProGlnArgHisLeuProValGlyCysValSerPheG 1648  
:|||||..... 1399 :|||||.....  
2529 AGCATGTGAGAGACCGGAATATCTCTCTGTGGATGGTATACATTTCA 2578  
1648 nAsnIleSerSerAsnValLeuGluGlnSerAlaIleSerAspAspIleL 1665  
|||||..... 1399 |||||.....  
2579 GAATATTTCTATATGTTTAAAGAAATGCGGCTCAGATGAGTGG 2628  
1665 euSerProAspGluGlyGlyPheCysSerGlyLysHisPheThrGluLeu 1681  
:|||||..... 1399 :|||||.....  
2629 TATCTCAGATTAAGAAGATATCTGCTCGAAATACTTTACTGATGCTCA 2678  
1682 GluLeuValGlyLeuLeuGluGlnAlaAlaGlyTyrPheThrMetGlyG 1698  
|||||..... 1399 |||||.....  
2679 GGACTTGTGGATTTACTGGAACAACAGCTGCTCTCTATGGCTGG 2728  
1698 yLeuTyrGluAlaValAsnGluValTyrLysAsnLeuIleProIleLeuG 1715

```

1729 CATGATTCAGACAGTTAATGAAGTTTACAAAGTACTTATTCCTATTCATG 2778
1729 ||:|||||
1715 luAlaHhISArGspTyrlYsLysLeuAlaAlaValhISGLyLysLeuGln 1731
1715 |||||
2779 AAGCTAAACGGGATGCCAAGAACTATCCCAATCATGTGTAACCTCAA 2828
1732 GluAlaPheThrLysIleMetHisGlnSerSerGlyTyrGluArgValPh 1748
1732 |||||
2829 GAACGATTCAGCAAAATTTGTCATCAGAGTACGCTGGGAGCGGAGCTT 2878
1748 eGlyThrTyrPheArgValGlyPheTyrGlyAlaHisPheGlyAspLeuA 1765
1748 |||||
2879 TGGACCTCATTTTCGTGTGTTTATGGAACCACTTCGGGAGCTTTGG 2928
1765 spGIUGInGluPheValTyrLysGluProSerIleThrLysLeuAlaGlu 1781
1765 |||||
2929 ATGAACAAGAATTTGTTTACAGAGAGCTGCATACCAAACTTGCAGAG 2978
1782 IleSerHisArgLeuGluGluPheTyrThrGluArgPheGlyAspArgVa 1798
1782 |||||
2979 ATATCTCAGAGATTGGAGGATTTTACGAGAAAGATTGGAGAGCATGT 3028
1798 lValGluIleLeysAspSerTyrProValAspLysSerLysLeuAspS 1815
1798 |||||
3029 GGTGAAGTATCAACAACTCTAATCCTGTAGACAGGTGTAATTTAGATC 3078
1815 eGlnLysAlaTyrIleGlnIleThrTyrValGluProTyrPheAspThr 1831
1815 |||||
3079 CTACACAGGATATATCAGATTACCTATGAGAGCCATCTTGCACACA 3128
1832 TyrgLueLysAspArgValThrTyrPheAspArgAsnTyrGlyLeuAr 1848
1832 |||||
3129 TATGAGATGGAAGACGAATCACCATTCTGACAAAATTTACATCTTGC 3178
1848 gThrPheLeuPheCysThrProPheThrProAspGlyArgAlaHisGlyG 1865
1848 |||||
3179 TCGATTATGATCTGTACACCTTTACTTACTTATGAGCCGCGCCATAGGG 3228
1865 lueuProGluGlnHisLysArgLysThrLeuLeuSerThrAspHisAla 1881
1865 |||||
3229 AACTTCATGAACAATTCAAAAGAGACCAATCTGACTACGTCATCAGGCC 3278
1882 PheProTyrIleLysThrArgIleArgValCysHisArgGluGluThrVa 1898
1882 |||||
3279 TTTCCTTAATTTAAACAAAGGCTCAATGCTACCTATTAAGAGAGATCA 3328
1898 lLeuThrProValGluValAlaIleGlnLysPheGlnLysLysThrArgG 1915
1898 |||||
3329 CTTAACACCAATTTGAAGTGTCTATGAGACATGACGAAAGAGACAGG 3378
1915 lueuAlaPheAlaThrGluGlnAspProProAspAlaLysMetLeuGln 1931
1915 |||||
3379 AGTTGGCATTTGCACACATCAGGATCCGACAGACCCCAAAATGCTTCAG 3428
1932 MetValLeuGlnGlySerValGlyProThrValAsnGlnGlyProLeuG 1948
1932 |||||
3429 ATGCTACTCCAGGATCTAGGACACACAGATGAATCAGGGGCTTTGGA 3478
1948 uValAlaGlnValPheLeuAlaGluIleProGluAspProLysLeuPheA 1965
1948 |||||
3479 AGTTGCCCGAGTTTCTGTGAAATACCTAGTAGCCCAAGAGCTCTTCA 3528
1965 rGHISHisAsnLysLeuArgLysPheLysAspPheCysLysLysCys 1981
1965 |||||
3529 GACATCATATATAAAGTCGACCTCTGTTAAAGATTTTACTAATAAGGTGT 3578
1982 GluAspAlaLeuArgLysAsnLysAlaLeuIleGlyProAspGlnLysG 1998
1982 |||||
3579 GAAAGATGCTTAAGAAAAATAAGAGCTTAATTTGGGCGGTTCAAAAGGA 3628
1998 uTyrHisArgGluLeuGluArgAsnTyrCysArgLeuArgGluAlaLeu 2014
1998 |||||

```

```

3629 GTATCAAGGGAATTC...GGGAAACTATCTTCGCTTAAAGAGGCCCTTA 3675
2015 Gln.ProLeuLeuThrGlnArgLeuProGlnLeuMetAlaProThr.... 2029
3676 CAGCCCT...AGATCACAAGAAAGTCCCTCAGTTATTCACAAGCAGATATGC 3722
2030 ..ProGluGluLeuArgAsnSerLeuAsnArgAlaSerPheArgLysAla 2045
3723 TGTCCCTCCGACACAGAGATTCCTTCAGTCGAATGAGCTTTCGCAAAATG 3772
2046 AspLeu 2047
3773 GATCTC 3778

seq_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH41911
seq_documentation_block:
ID AAH41911 standard; cDNA: 4143 BP.
XX
AC AAH41911;
XX
DT 30-AUG-2001 (first entry)
XX
DE Preliminary human CLASP-3 cDNA sequence Fig 1.
XX
KW Human; cadherin-like asymmetry protein; CLASP; CLASP-3; immune response;
KW cell surface molecule; transmembrane protein; immunosuppressive; vaccine;
KW antiinflammatory; antiarthritic; antianemic; dermatological; uropathic;
KW ophthalmological; antineumatic; nephrotropic; antidiabetic;
KW neuroprotective; antislumatic; antibacterial; antilensense therapy;
KW gene therapy; chromosome 1p31.1; ss.
XX
OS Homo sapiens.
XX
PN W0200142297-A2.
PD 14-JUN-2001.
XX
PE 13-DEC-2000; 2000MO-US34171.
XX
PR 13-DEC-1999; 99US-0170453.
PR 14-JAN-2000; 2000US-0176195.
PR 14-FEB-2000; 2000US-0182296.
PR 11-APR-2000; 2000US-0196267.
PR 11-APR-2000; 2000US-0196460.
PR 11-APR-2000; 2000US-0196527.
PR 11-APR-2000; 2000US-0196528.
PR 11-APR-2000; 2000US-0547276.
PR 13-OCT-2000; 2000US-0240503.
PR 13-OCT-2000; 2000US-0240508.
XX
PA (ARBO-) ARBOR VITA CORP.
XX
PI Lu P, Garman JD, Candia AF;
XX
DR WPI; 2001-375003/39.
XX
DR P-PADB; AAB99493.
XX
PT Novel isolated cadherin-like asymmetry protein (CLASP) -3, useful for
PT useful in preventing or treating a CLASP-3-mediated disease e.g.
PT autoimmune disease
XX
PS Example 5; Fig 1; 189pp; English.
XX
XX
The present invention describes an isolated polypeptide (II) comprising a
nucleotide (nt) sequence (Sta) that has at least 90% identity to the
sequence given in AAH41934 and is immunologically cross-reactive with
the derived amino acid (aa) sequence (SId) given in AAB99495 or shares a
biological function with native CLASP-3. (I) has immunosuppressive,
antiinflammatory, antiarthritic, antianemic, dermatological, uropathic,
CC ophthalmological, antineumatic, nephrotropic, antidiabetic,
CC neuroprotective, antislumatic and antibacterial activities, and can be
used in antilensense therapy, vaccine production and gene therapy. CLASP-3

```



CC related sequences can be used in preventing or treating a CIASP-3 mediated disease, preferably an autoimmune disease by inhibiting an CC immune response. The autoimmune disease is caused or exacerbated by CC increased activity of T helper 1 lymphocytes (TH1). AAH41893 to AAH41953 CC and AA99491 to AA99507 represent sequences which are used in the CC exemplification of the present invention. CIASP-3 is localised in the CC chromosome location 1p31.1.

XX Sequence 4143 BP; 1251 A; 829 C; 890 G; 1173 T; 0 other;

## alignment\_scores:

Quality: 4273.00 Length: 1247  
Ratio: 4.039 Gaps: 11  
Percent Similarity: 84.844 Percent Identity: 67.121

## alignment\_block:

US-09-736-968A-2 x AAH41911 ..

Align seg 1/1 to: AAH41911 from: 1 to: 4143

```
877 ArgSerLysSerIleSerSerSerSerProAspLeuAlaValAlaProG1 893
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10 CGTTCGAGAGCCTTAGTATATACCATCCAGATATATCTGGGACTCCAC 59
893 ySerValAspAspGluValSerArgIleLeuAlaSer..... 905
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 GTCACGAGATGATGAGTTCGATCATCATCGGAGTAAAGGTTGATGATC 109
905 ..... 905
110 GCTCCATTCTCTGGTTCACACTGTGTCCAAAGCTGCCCATGGGA 159
905 ..... 905
160 TCCACCCAGTCCAGTGCAGATCAACACAGCTATGATGAGATG 209
905 ..... 905
210 TAATCGTAGTCTTCCACACAGAGCGTCAAGTTCTTACAAACATTA 259
906 ..... 905
260 CGGAGCGCTTACCACTAAAGCTTTTTCACGAGAGCTGGCTTTCAG 309
916 TrpValValSerSerSerAlaValArgIleAlaIleLeuGlnHisAlaTr 932
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
310 TGGGTGTTGTCAGTGCAGCGTTCGGAATCAGCTTGGCAACAGCCTG 359
932 pPhePhePheGlnLeuMetValLysSerMetAlaLeuHisLeuLeuG 949
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 GTTCCTTTTGAATTAATGTTAAAGAGCATGTCACATTTACTTTA 409
949 LysGlnArgLeuAspTrpProArgLysLeuArgPheProGlyArgPheLeu 965
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
410 ATGATTAACCTTGAGGCTCCAGAGAAAAGTCTTTTCCAGACGCTTTCATG 459
966 AspAspIleThrAlaLeuValGlySerValGlyLeuGlnValIleThrAr 982
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
460 GATGACATTCAGCTCTGTTCACACAGATGCTAGTATAGTATGTTCAAG 509
982 gValHisLysAspValGluLeuAlaGlnHisLeuAsnAlaSerLeuAlaP 999
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
510 ATTTCAGAAAGACACAGAAATGTTGAGAGACTCAATACAGCCTTGCAAT 559
999 hePheLeuSerAspLeuSerLeuValAspArgGlyPheValPheSer 1015
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
560 TCTTTCATATGATCTGTTCTCTATGACAGAGATTTGTTTATAGC 609
1016 LeuValArgAlaHisTyrLysGlnValAlaThrArgLeuGlnSerPerP 1032
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
610 CTATATAAGTCTGCTATATAACAGGTGCTTCAAGCTTACTACTTACC 659
```

```
1032 CAsnProAlaAlaLeuLeuThrLeuArgMetGluPheThrArgIleLeuG 1049
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
660 GAATCCAGCTTCTGAGTGTCTGAGGCTGATTTCTTCAAGATCAATCT 709
1049 ySerHisGlnHisTyrValThrLeuAsnLeuProCysCysProLeuSer 1065
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
710 GCAGTCATGAGCAGCATGTATACATTAACTTACCTGAGCTTACTTACT 759
1066 ProProAlaSerProSerProSerValSerSerThrThrSerGlnSerSe 1082
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
760 CCACCTGCATCTCCATCACCCTTCTGTTCTTCTGCAACATCTCAGAGTTC 809
1082 rThrPheSerSerGlnAlaProAspProLysValThrSerMetHegUL 1099
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
810 TGGATTTCTACGAATGTACAAAGAACAAAAGATATGTTGTTGAT 859
1099 eUeSerGlyProPheArgGlnGlnHisPheLeuAlaGlyLeuLeuThr 1115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
860 TATCCGTCCTTTCGCCACACAGCATTTTGGCAGGACTGTGTATACA 909
1116 GluLeuAlaLeuAlaLeuGluProGluAlaGluGlyAlaPheLeuLeuH 1132
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
910 GAGCTGGCTGTCATTTTAGACCTGATGCTGAGGACTGTTGGATTGCA 959
1132 sLysLysAlaIleSerAlaValHisSerLeuLeuGlyHisAspThrA 1149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
960 TAAAGAACTCATCAATATGATGTACACAAATTTACTCTCCGTCACAGCTCAG 1009
1149 sPProArgTyrAlaGluAlaThrValLysAlaArgValAlaGluLeuTyr 1165
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1010 ACCCGGCTACTCTGACCTCAGATTAAGGCTCGAGTGGCATGTGTAT 1059
1166 LeuProLeuLeuSerIleAlaArgAspThrLeuProArgLeuHisAspPh 1182
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1060 CTACCTGTGATGTATATCATGGAACCTGATCCTCAGTGTATGATTT 1109
1182 eAlaGluLeuGlyProGlyGlnArgSerArgLeuAlaSerMetLeuAspSera 1199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1110 TACAGAACTCACAATCAACAGAGGAAACCAATTTGTATGCCACTGTG 1159
1199 sPThrGluGlyGluLysAspIleAlaGlyThrIleAsnProSerValAla 1215
|| |||:||||:||||:||||:||||:||||:||||:||||:
1160 ATTATGAAGATGAGAGC.....GGAAGTATGATTAAGCAACCGTTGCC 1203
1216 MetAlaIleAlaGlyGlyProLeu.....AlaProGlySerAr 1228
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1204 ATGGCAATCGCAGGAGACATGCGTCCCTCACTAACAGAGCCTGCAAGTT 1253
1228 gAlaSerIleSerGlnGlyProProThrAlaSerArgAlaGlySAlaL 1245
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1254 CCTCTCAGCTCA.....ACGAGTGGCAGGCAACACACTACTCT 1291
1245 eUeSerAlaGluSerSerArgThrLeuAlaCysValLeuTrpValLeu 1261
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1292 TTTTCAGCAAGATCAAGTGCAGAGCCTTTTGATCTGCTACTTGGGTTCTC 1341
1262 LysAsnThrGluProAlaLeuLeuGlnArgTrpAlaThrAspLeuThrLe 1278
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1342 AAAAATGACAGATGAAGACAGTCTTACAGAAAGTGTATACAGATCTCAGT 1391
1278 uProGlnLeuGlyArgLeuLeuAspLeuLeuTyrLeuGlySAlaAlaP 1295
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1392 CTTCGAGCTAAACCGGCTATTAGATCTCTTATCTCTGTGTGCTTGGCT 1441
1295 heGlyTyrLysGlyLysLysAlaPheGluArgIleAsnSerLeuThrPhe 1311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1442 TTGATATATAAAGGAAAAAGTGTGAAACGATGAATAGCTTGACCTTT 1491
1312 LysLysSerLeuAspMetLysAlaArgLeuGluGluAlaIleLeuGlyTh 1338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1492 AAGAAATCAAAAAGACATGAGCAAAAGCTTGAGAGACTATTTCTGGAG 1541
1338 rLleGlyAlaArgGlnGluMetValArgArgSerArg.....GluA 1342
```



```
1938 a1g1yProthValAsnGlnGlyProLeuGluValAlaGlnValPheLeu 1954
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
3392 TAGGACACAGATGATCATCGGGGCTTTGGAACTTGCACAGCTTTTCTG 3441
1955 AlaGluLeuProGluAspProValLeuPheArgHisAsnLysLeuAr 1971
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
3442 TCTGAAATACCTAGTACGACCCAAAGCTCTTCAGACATCATTAATACTGGC 3491
1971 gLeuGlyPheLysAspPheCysLysCysGluAspAlaLeuArgLysA 1988
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
3492 ACCTGCTTTAAAGATTTTACTAAAGGTGTGAAGATGCGCTTAGAAAA 3541
1988 snLysAlaLeuIleGlyProAspGlnLysGluTyrHisArgGluLeuGln 2004
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
3542 ATAAAGAGCTTAATTGGCCGCTTCAAAAGAGTATCAAGGGAATTC... 3588
2005 Arg AsnTyrCysArgLeuArgGluAlaLeuGln ProLeuLeuThrGln 2020
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
3589 GGGAAACATATCTTGGCTTAAAGAGGCCCTTACAGCCCT...AGATCACAG 3635
2021 ArgLeuProGlnLeuMetAlaProThr.....ProProGlyLeuArgAs 2035
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
3636 AAAGTCCCTCAGTTATCCAAAGCCAGTATGTGCTTGTCCCTGGCCACAGAGA 3685
2035 nSerLeuAsnArgLasePheArgLysAlaAspLeu 2047
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
3686 TTCCTTCAGTCGAACGAGCTTTCGAAATGATGATCTC 3722
seq_name: /SID1/gcgdata/geneseq-emb1/NA2001A.DAT:AAH41912
seq_documentation_block:
ID AAH41912 standard; cDNA: 4144 BP.
XX
AC AAH41912;
XX
DT 30-AUG-2001 (first entry)
XX
DE Human CLASP-3 cDNA sequence Fig 4.
XX
KW Human, cadherin-like asymmetry protein; CLASP; CLASP-3; immune response;
cell surface molecule; transmembrane protein; immunosuppressive; vaccine;
antiinflammatory; antiarthritic; antianaemic; dermatological; uropathic;
ophthalmological; antinephrotic; nephrotropic; antidiabetic;
KW neuroprotective; antinephrotic; antidiabetic; antidiabetic;
KW gene therapy; chromosome 1p31.1; ss.
XX
OS Homo sapiens.
XX
PN WO200142297-A2.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WC-US34171.
XX
PR 13-DEC-1999; 99US-0170453.
PR 14-JAN-2000; 2000US-0176195.
PR 14-FEB-2000; 2000US-0182296.
PR 11-APR-2000; 2000US-0196267.
PR 11-APR-2000; 2000US-0196460.
PR 11-APR-2000; 2000US-0196527.
PR 11-APR-2000; 2000US-0196528.
PR 11-APR-2000; 2000US-0547276.
PR 13-OCT-2000; 2000US-0240503.
PR 13-OCT-2000; 2000US-0240508.
XX
PA (ARBO-) ARBOR VITA CORP.
XX
PI Lu P, Garman JD, Candia AF;
XX
DR WPI: 2001-375003/39.
DR P-PSDB; AAB99494.
XX
```

```
PT Novel isolated cadherin-like asymmetry protein (CLASP) -3, useful for
PT useful in preventing or treating a CLASP-3-mediated disease e.g.
PT autoimmune disease
XX
PS Disclosure: Fig 4A; 189pp; English.
XX
CC The present invention describes an isolated polypeptide (I) comprising a
CC nucleotide (nt) sequence (S1a) that has at least 90% identity to the
CC sequence given in AAH41934 and is immunologically cross-reactive with
CC the derived amino acid (aa) sequence (S1b) given in AAB99495 or shares a
CC biological function with native CLASP-3. (I) has immunosuppressive,
CC antiinflammatory, antirheumatic, antianaemic, dermatological, uropathic,
CC ophthalmological, antinephrotic, nephrotropic, antidiabetic,
CC neuroprotective, antinephrotic and antibacterial activities, and can be
CC used in antisense therapy, vaccine production and gene therapy. CLASP-3
CC related sequences can be used in preventing or treating a CLASP-3
CC mediated disease, preferably an autoimmune disease by inhibiting an
CC immune response. The autoimmune disease is caused or exacerbated by
CC increased activity of T helper 1 lymphocytes (TH1). AAH41893 to AAH41953
CC and AAB99491 to AAB99507 represent sequences which are used in the
CC exemplification of the present invention. CLASP-3 is localised in the
CC chromosome location 1p31.1.
XX
SQ Sequence 4144 BP; 1252 A; 829 C; 890 G; 1173 T; 0 other:

alignment_scores:
Quality: 4273.00 Length: 1247
Ratio: 4.039 Gaps: 11
Percent Similarity: 84.844 Percent Identity: 67.121

alignment_block:
US-09-736-968A-2 x AAH41912 ..
Align seg 1/1 to: AAH41912 from: 1 to: 4144

877 ArgSerLysSerLysSerSerSerAsnProAspLeuAlaValAlaProGln 893
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
11 GGTTCGAGAGCCTTAGATATAGCAATCCAGATATCTGGAGCCTCCAC 60
893 ySerValAspAspGluValSerArgIleLeuAlaSer..... 905
||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
61 GTCACACAGATGATGAAGTTCGATCATCATCGGAGCTTAAGGTTAGATC 110
905 ..... 905
111 GCTCCAAATCTGTGGTAAACACTGTGTCTCAAAAGCTGCCCATGGGA 160
905 ..... 905
161 TCCAAACCCAGTCCAGTGCAGATCAACACAGGCTATGATCGAATTG 210
905 ..... 905
211 TAATGTAATGTCTTGCACACAGACGCTCAAGTTCTTCAACACATTAA 260
906 ..... 906
261 CGGAGCGCTTACCACTAAAGAGCTTTTCAAGAGAGCTGGCTTTCGAG 310
916 TTPValAlaSerSerSerAlaValAlaArgLysAlaLeuGlnHisAlaTr 932
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
311 TGGGTGTGTTGCGAGTGGCAGCGCTTGGGAATCAGCTTTGCAACAAGCTG 360
932 PhePhePheGlnLeuMetValLysSerMetAlaLeuHisLeuLeuG 949
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
361 GTTCTTTTGAATTAATGTAAGAGCATGTGACCATTTATACCTTA 410
949 LysGlnArgLeuAspThrProArgLysLeuArgPheProGlyArgPheLeu 965
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
411 AGATTAACCTGAGGCTCCAGGAAAGTCTTTCCAGAGCTTCATG 460
966 AspAspIleThrAlaLeuValGlySerValGlyLeuGluValIleThrAr 982
```



```

1571 sGInGluAspProGluMetLeuIleAspLeuMetTyrArgIleAlaArg 1588
|||||
2293 CCAGGAGGATCCGAAATGTGATGCTAAATGACAAATGGCCAAAG 2342
1588 lYTrGInGlySerProAspLeuArgLeuThrTriPleuGlnAsnMetAla 1604
|||||
2343 GTTACACAGACCTCTCCAGATCTGCGATGACCTGGTGGCAAGCATGGCA 2392
1605 gLYlYsHsAlaGluDeuGlnYsHsAlaGluAlaAlaGlnCysMetVa 1621
|||||
2393 GCGAAGCAGCTCAGAACGAAATCATGCTGAAGCTGCACAGTGTCTACT 2442
1621 lHsAlaAlaAlaAlaLeuValAlaGluTyrIleuAlaLeuLeuGluAspGlnA 1638
|||||
2443 CCAGCTGAGAGCAGCTGTGCTGAATATTGAGCATGCGAGGAGCCGGA 2492
1638 rGHisLeuProValIleYlYsValSerPheGlnAsnIleSerAsnVal 1654
|||||
2493 AATATCTCTCTGSGSAGTGTACATTTCAGAAATATTTCATCTAATGTT 2542
1655 LeuGluGluSerAlaIleSerAspAspIleLeuSerProAspGluGlu 1671
|||||
2543 TTGAGAGAAATCTGCGTCTCAGATGATGTATCTCCAGATGAAGAAAG 2592
1671 yPheCysSerGlyYsHsPheThrGluLeuGlyLeuValIleGlyLeuG 1688
|||||
2593 TATCTGCTCTGGAAATACTTACTGAGTCAGAGACTGTGGGATTAATCG 2642
1688 lUGlnAlaAlaGlyYrPheThrMetGlyIleuTyrGluAlaValAsn 1704
|||||
2643 AACAAACAGCTGCTCTCTCTATGCGTGCATGTATGAAGCAGTATAT 2692
1705 GluValTyrYlYsAsnLeuIleProIleLeuGluAlaHsArgAspTyr 1721
|||||
2693 GAAGTTTAAAGTAACTTATCTTATTCATTAAGAAATCGGGATGCCAA 2742
1721 sLYsLeuAlaAlaValHsGlyYsLeuGlnGluIlePheThrIleM 1738
|||||
2743 GAACATATCCACATATCATGTAACCTCAAGAAAGCATTCACAAATGG 2792
1738 eThiGlnSerSerIlyTyrGluArgValIlePheGlyThrTyrPheArgVal 1754
|||||
2793 TTCATCAGAGTACTGCTGCGAGCGGATGTTGGCACCTATTTCGTCT 2842
1755 GlyPheTyrGlyAlaHsPheGlyAspLeuAspGluGlnPheValTyr 1771
|||||
2843 GGTTTTATGGAACCAAGTTCGGGATTTGGATGAACAAGAAATTTGTTA 2892
1771 rLYSGluProSerIleThrYsLeuAlaGluIleSerHisArgLeuGluG 1788
|||||
2893 CAAGGAGCTGCGAATACCAAACTTGCAGAGATCTCTCAAGATTGGAGG 2942
1788 lUPheTyrThrGluArgPheGlyAspAspValValGluIleIleuYsAsp 1804
|||||
2943 GATTTTACGAGAAAGATTGGAGAGATGCTGTTGAATTAATCAAGAC 2992
1805 SerTyrProValAspYsSerYsLeuAspSerGlnYsAlaTyrIleG 1821
|||||
2993 TCTAAACCTGTAGACCAAGTGAATAGATCTCAACAAAGCATTAATCA 3042
1821 nIleThrTyrValGluProTyrPheAspThrTyrGluLeuYsAspArgV 1838
|||||
3043 GATTACCTATGTGAGCCATACTTGTACACATATGATGATGAAGGACAGA 3092
1838 aLThrTyrPheAspArgAsnTyrGlyLeuArgThrPheLeuPheCysThr 1854
|||||
3093 TCACCTATTTCGCAAAATTAACAATCTCGTCGATTCATGTACGTACGA 3142
1855 ProPheThrProAspGlyArgAlaHsGlyGluLeuProGluGlnHisLy 1871
|||||
3143 CCTTACTTATGATGGCCGTGCCATGGGGAACCTTCATGAACAATTCGA 3192

```

```

1871 sArgLYsThrIleuLeuSerThrAspHisAlaPheProTyrIleYsThrA 1888
|||||
3193 AAGGAAGACCATCTCTGACTACCTCATGCTTCCTTATATATTAACAA 3242
1888 rGlyIleArgValCysHsArgGluThrValIleuThrProValGluVal 1904
|||||
3243 GGGTCATATGCTACTATMAAGAAAGAGATCATCTTAACACCAATGAAGTT 3292
1905 AlaIleGluAspMetGlnLYsThrArgGluLeuValAlaPheAlaThrG 1921
|||||
3293 GCTATTGAGGACATGCGCAAAAAGACACAGAGATTGGCATTTCGAACACA 3342
1921 uGlnAspProProAspAlaLYsMetLeuGlnMetValLeuGlnGlySerV 1938
|||||
3343 TCAGATATCCCGCAGACCCCAAAATGCTTCAGATGTACTCCAGGAGACTG 3392
1938 aLGlyProThrValAsnGlnGlyProLeuGluValAlaGlnValPheLeu 1954
|||||
3393 TAGGCACACAGTGNATCAGGGGCTTTGGAAAGTTGCCAGGTTTCTCTG 3442
1955 AlaGluIleProGluAspProLYsLeuPheArgHisAsnLYsLeuAr 1971
|||||
3443 TCTGAATATCACTAGTACCCAAAGCTCTTCAGACATCATATAAATGCG 3492
1971 gLeuCYsPheLYsAspPheCysLYsCYsGluAspAlaLeuArgLYsA 1988
|||||
3493 ACTCTGCTTAAAGATTTTACTTAAAGGTGTGAAGATGCTTAAAGAAAA 3542
1988 sLYsAlaLeuIleGlyProAspGlnLYsGluTyrHisArgGluLeuGlu 2004
|||||
3543 ATAAGAGCTTATATTGGCGCGTTCAAAAAGAGTATCAAAAGGAAATG... 3589
2005 Arg.AsnTyrCysArgLeuArgGluAlaLeuGln.ProLeuLeuThrGln 2020
|||||
3590 GCGAAACATATCTTCCCTTAAAGAGCCCTACAGCCCT...AGATCACAG 3636
2021 ArgLeuProGlnIleuMetAlaProThr.....ProProGlyLeuArgAs 2035
|||||
3637 AAGATCCCTCGATTATCCAGCCAGTATGCTTCCCTGCGCCACAGAGA 3686
2035 nSerLeuAsnArgAlaSerPheArgLYsAlaAspLeu 2047
|||||
3687 TTCCTTCACTGGAATAGACCTTTCGCAAAATGATCTC 3723
seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.AAH41952
seq_documentation_block:
ID AAH41952 standard; DNA; 3173 BP.
XX
AC AAH41952:
XX
DT 30-AUG-2001 (first entry)
XX
DE CLASP-3 nucleotide fragment.
XX
KW Human; cadherin-like asymmetry protein; CLASP; CLASP-3; immune response;
KW cell surface molecule; transmembrane protein; immunosuppressive; vaccine;
KW antiinflammatory; antitumor; antianemic; dermatological; utropathic;
KW ophthalmological; antirheumatic; nephrotoxic; antithyroid; antidiabetic;
KW neuroprotective; antistatic; antibacterial; antisense therapy;
KW gene therapy; chromosome 1p31.1; ds.
XX
OS unidentified.
XX
PN WO200142297-A2.
XX
PD 14-JUN-2001.
XX
PE 13-DEC-2000; 2000MO-US34171.
XX
PR 13-DEC-1999; 99US-0170453.
PR 14-JAN-2000; 2000US-0176195.
PR 14-FEB-2000; 2000US-0182296.

```

PR 11-APR-2000; 2000US-0196267.  
PR 11-APR-2000; 2000US-0196460.  
PR 11-APR-2000; 2000US-0196527.  
PR 11-APR-2000; 2000US-0196528.  
PR 11-APR-2000; 2000US-0547276.  
PR 13-OCT-2000; 2000US-0240503.  
PR 13-OCT-2000; 2000US-0240508.  
XX  
XX (ARBO-) ARBOR VITA CORP.  
XX  
XX Lu P, Garman JD, Candia AF;  
XX WPI: 2001-375003/39.  
XX  
XX Novel isolated cadherin-like asymmetry protein (CLASP) -3, useful for  
XX PT useful in preventing or treating a CLASP-3-mediated disease e.g.  
XX PT autoimmune disease -  
XX  
XX Disclosure: Fig 9A: 189bp; English.  
XX  
XX The present invention describes an isolated polypeptide (I) comprising a  
XX CC nucleotide (nt) sequence (sta) that has at least 90% identity to the  
XX CC sequence given in AAH41934 and is immunologically cross-reactive with  
XX CC the derived amino acid (aa) sequence (S1b) given in AAB99495 or shares a  
XX CC biological function with native CLASP-3. (I) has immunosuppressive,  
XX CC antiinflammatory, antirheumatic, antianemic, dermatological, urepatic,  
XX CC ophthalmological, antithrombotic, nephrotropic, antithyroid, antidiabetic,  
XX CC neuroprotective, antistatic and antibacterial activities, and can be  
XX CC used in antisense therapy, vaccine production and gene therapy. CLASP-3  
XX CC related sequences can be used in preventing or treating a CLASP-3  
XX CC mediated disease, preferably an autoimmune disease by inhibiting an  
XX CC immune response. The autoimmune disease is caused or exacerbated by  
XX CC increased activity of T helper 1 lymphocytes (TH1). AAH41933 to AAH41953  
XX CC and AAB99491 to AAB99507 represent sequences which are used in the  
XX CC exemplification of the present invention. CLASP-3 is localised in the  
XX CC chromosome location 1p31.1.  
XX  
XX Sequence 3173 BP; 981 A; 624 C; 688 G; 880 T; 0 other;  
SQ

alignment\_scores:  
Quality: 3487.50 Length: 952  
Ratio: 4.127 Gaps: 10  
Percent Similarity: 88.761 Percent Identity: 71.744

alignment\_block:  
US-09-736-968a-2 x AAH41952 ..

Align seg 1/1 to: AAH41952 from: 1 to: 3173

1111 GlyLeuLeuLeuThrGluLeuAlaLeuAlaLeuGluProGluAlaGlu 1127  
4 GGACTTGTGTTTCAGAGCTGGCTGTCATTGTTAGACCTGATCTGAGG 53  
1127 yAlaPheLeuLeuHisLysLysAlaIleSerAlaValHisSerLeuLeu 1144  
1111  
54 ACCTGTTGATGTCATGAAGAAAGTCATCATATGTCACACATTTACTCT 103  
1144 ySgLIHisAspThrAspProArgTyrIaGluAlaThrValLysAlaArg 1160  
104 CCACTCACTGACTGACACCGCGGTACTGTGACCTCGATTAAGGCTCGA 153  
1161 ValAlaGluLeuTyrLeuProLeuLeuSerIleAlaArgAspThrLeu 1177  
1111  
154 GTGGCCATGTTGATCTACCTGATGTTATTCATGGAAGTGTAC 203  
1177 cATGLeuHisAspPheAlaGluGlyProGlyGlnArgSerArgLeuAla 1194  
204 TCAGCTGTATGATTTACAGAACTCAACATCAACGAGAGCAACTTT 253  
1194 eMetLeuAspSerAspThrGluGlyGluGlyAspIleAlaGlyThrIle 1210  
254 GTATAGCCACTGATGATTAAGAAAGTGAAGC.....GGAAGTATGATA 297

1211 AsnProSerValAlaMetAlaIleAlaGlyProLeu..... 1223  
1224 AlaProGlySerArgAlaSerIleSerGlnGlyProThrAlaSer 1240  
348 AAGCCCTGGCAGTTTCCCTCCACGTC.....ACGAGTGCA 385  
1240 rGAlaGlyCysAlaLeuSerAlaGluSerSerArgThrLeuLeuAlaCys 1256  
386 GGCACACACTTACTCTTTCAGCAGAAATCAAGTCAGAGCTTTTATCTGT 435  
1257 ValLeuTrpValLeuLysAsnThrGluProAlaLeuGlnArgTrpAl 1273  
436 CTACTTTGGGTTTCCAAAATGACAGATGAACACTTTCACGAGAGTGT 485  
1273 aThrAspLeuThrLeuProGlnLeuGlyArgLeuLeuAspLeuTyrL 1290  
486 TACAGATCTCTCAGCTTTCAGCTAAACCCGCTATTGATCTGCTTATC 535  
1290 eCysLeuAlaAlaPheGluTyrLysGlyLysAlaPheGluArgLe 1306  
536 TCTGTGTCTTCTGCTTGTAGATTAAGGAAAGAGTTTGAACGAAATG 585  
1307 AsnSerLeuThrPheLysLysSerLeuAspMetLysAlaArgLeuGlu 1323  
586 AATAGCTTGACCTTTAAGAAATCAAAAGACTAGAGAGCAAGCTTGAAGA 635  
1323 uAlaIleLeuGlyThrIleGlyAlaArgGlnGluMetValArgArgSer 1340  
636 AGCATTTCTTGGGAGCATAGTGCAGGCAAGAAATGTTAGCGGCAAGCC 685  
1340 rG.....GluArgSerPro.....PheGlyAsnProGlu 1349  
686 GAGCAGACCTCGACAGAAACCCATCTGGAAGCTTTGGAAAGTCAGAA 735  
1350 AsnValArgTrpArgLysSerValThrHisTrpLysGlnThrSerAsp 1366  
736 AATTGAGGTGAGAGAAAGATATGACTCAGCGGTCAAAACACAGAGAA 785  
1366 gValAspLysThrLysAspGluMetGlnHisGluAlaLeuValGluGly 1383  
786 GCTTGACAAATCAAGACAGAGATTGAACACGAAAGCATTTCTGTGGAGT 835  
1383 snLeuAlaThrGluAlaSerLeuValValLeuAspThrLeuGluIle 1399  
836 ACCTGGCTACAGAAAGCAAGCTATCATTTAGATTTAGATTTGATGTT 885  
1400 ValGlnThrValMetLeuSerGluAlaArgGluSerValLeuGlyAla 1416  
886 GTTCAGACCCGTTCTGTACAGCAATCCAAAGAGCATTTCTGTGGAGT 935  
1416 IleuLysValValLeuTyrSerLeuGlySerAlaGlnSerIleLeuPhe 1433  
936 GCTAAAGTGTCTACTACACAGCATGGCTGTAAACAAAGTGCAGTTATC 985  
1433 eGlnHisGlyLeuAlaThrGlnArgAlaLeuValSerLysPheProGlu 1449  
986 TACAACACTGTTTGTACACAGAGAGCTTGTTCAAAGTTCCGGA 1035  
1450 LeuLeuPheGluGluAspThrGluLeuCysAlaAspLeuCysLeuArg 1466  
1036 CTCTTATTTGAAGAAGACAGACAGTGTGCTGATTTATGCTCAGAGCT 1085  
1466 uLeuArgHisCysGlySerArgIleSerThrIleArgThrHisAlaSer 1483  
1086 TCTCCGACACTGTAGACCATCGGTACAAATTCGGTCAACCCCGTGT 1135  
1483 IAspLeuTyrLeuLeuMetArgGlnAsnPheGluIleGlyHisAsnPhe 1499  
1136 CTTCCTTTTACTTACTATATGAGCAAACTTTGAGATTGGGAATTAAC 1185

```

1500 AlaArgValLysMetGlnValThrMetSerLeuSerLeuValGlyTh 1516
|||||
1186 GCCAGGTTAAATGACAGGTACCAATGTCACTATCCCTGGGGGCAC 1235
1516 rThnGlnAspHisSerGlnGlnHisLeuArgArgSerLeuLysThrLeu 1533
|||||
1236 ATCTCAGAAATTTATAGAAATCTTAAGACGTTCTTAAGACTATAT 1285
1533 eutThrTyrAlaGlnGlnLysMetGlyLeuArgAspSerThrPheAlaGln 1549
|||||
1286 TGACATATGCTGGAAGAGATCTGSAATTTGAGGGAACAACATTTCCGAT 1335
1550 GlnValGlnAspLeuMetPheAsnLeuHisMetLeuLeuThrAspThrVa 1566
|||||
1336 CAGGTCACGAGATCGTGTTCATATCTCATATGATTTCTTCTGATACTGT 1385
1566 LysMetLysGlnHisGlnGlnLysAspProGlnMetLeuLeuLeuMet 1583
|||||
1386 GAAATGGAAGACACCCAGAGAGATCTGAAATGTGATGTATTAATGT 1435
1583 TyrArgLeuAlaArgGlyTyrGlnGlnLysSerProAspLeuArgLeuThrTrp 1599
|||||
1436 ACAGATTTGCCAAGAGGTACCAACCTCTCCAGATCTGCGATGACCTGG 1485
1600 LeuGlnAsnMetAlaGlyLysHisAlaGlnLeuGlnLysAsnHisAlaGln 1616
|||||
1486 TTGCAGAACATGGTAGCGACACTCAGACGAAAGCAATCATGCTGAAGC 1535
1616 AlaGlnLysMetValHisAlaAlaAlaLeuValAlaGlnTyrLeuAla 1633
|||||
1536 TGCAACAGTGTCTATGTCACACGACACTGTGTGCTGAATATTTGACCA 1585
1633 euLeuGlnAspGlnArgHisLeuProValGlyCysValSerPheGlnAsn 1649
|||||
1586 TGCTGGAGACCGZAAATATCTCTCTCCATGTTCAATTTGACAAT 1635
1650 LLeSerSerAsnValLeuGlnGlnLysSerAlaLLeSerAspAlaLeuSe 1666
|||||
1636 ATTTATCTTAATGTTTGAAGAAATCTCGGCTCTCATGATGATGGTATC 1685
1666 rProAspGlnGlnLysPheCysSerGlyLysHisPheThrGlnLeuGln 1683
|||||
1686 TCCAGATGAGAAAGTATCTGCTCTGGAATAATCTTACTGAGTCAGGAC 1735
1683 euValGlyLeuLeuGlnGlnAlaAlaGlyTyrPheThrMetGlyLysLeu 1699
|||||
1736 TTGTGGGATTTACTGGAACAAGCAGCTCTCTCTCTATGAGCTGCATG 1785
1700 TyrGlnAlaValAsnGlnValTyrLysAsnLeuLLeProLLeuGlnAla 1716
|||||
1786 TATGAGGCGTTAATGAACTTACAAAGTACTTATTCATTATCATGAAGC 1835
1716 HisArgAspTyrLysLysLeuAlaAlaValHisGlyLysLeuGlnGln 1733
|||||
1836 TTAATCGGATGCAAAAGAACTATCCACAATTTGATAAAGCTTCAAGAG 1885
1733 LAspThrLysLLeMetHisGlnSerSerGlyTyrPGLuArgValPheGly 1749
|||||
1886 CATTCAGCAAAATTTGTTCAACAGACTGCTCGGAGCGGAGTGTGGC 1935
1750 ThrTyrPheArgValGlyPheTyrGlnAlaHisPheGlyAspLeuAsp 1766
|||||
1936 ACCATATTTTCGTGTGTTTATGGAACCAAGTTCGGGGATTTGGATGA 1985
1766 uGlnGlnPheValTyrLysGlnProSerLLeThrLysLeuAlaGlnLys 1783
|||||
1986 ACAAGAAATTTGTTTACAGAGACCTGCATTAACCAAACTTGGAGATAT 2035
1783 eThHisArgLeuGlnGlnLysPheThrGlnArgPheGlyAspAspValVal 1799
|||||
2036 CTCACAGATTTGGAGGATTTTACGGGAAGAAATTTGGAAGAGATGGGTT 2085
1800 GlnLLeLLeLysAspSerTyrProValAspLysSerLysLeuAspSer 1816

```

```

|||||
2086 GAAGTATATAAAGACTTATATCTGTAGACAGGTAAATTTGATCCCTAA 2135
1816 nuLysAlaTyrLLeGlnLLeThrTyrValGlnProTyrPheAspThrTyrG 1833
|||||
2136 CAAGCATATATTCACATATACCTATGTGAGCCATACTTTGACACATAG 2185
1833 LuLeuLysAspArgValThrTyrPheAspArgAsnTyrGlyLeuArgThr 1849
|||||
2186 AGATGAAGACAGAAATCACCTATTTTGACAAAAAATTACATATCTGTGCA 2235
1850 PheLeuPheCysThrProPheThrProAspGlyArgAlaHisGlyLys 1866
|||||
2236 TTATGTACTGCTACACCTTTACTTATAGATGCCCGGAGGAGGAACT 2285
1866 uProGlnGlnHisLysArgLysThrLeuLeuSerThrAspHisAlaPhe 1883
|||||
2286 TCATGAACAATTCAAAAGAACACATCTGACTACGCTCATGCTTCG 2335
1883 rOtyrLLeLysThrArgLLeArgValCysHisArgGlnGlnLysThrVal 1899
|||||
2336 CTATATTTAAACAAAGGCTCAATGCTCATTAAGAAAGAGATCATCTTA 2385
1900 ThrProValGlnValAlaLLeGlnAspMetGlnLysLysThrArgGln 1916
|||||
2386 ACACCAATTAAGTTGCTATTTAGGACATGCAGAAAAAGACAGAGATT 2435
1916 uAlaPheAlaThrGlnGlnAspProProAspAlaLysMetLeuGlnMet 1933
|||||
2436 GGCATTTGCAACACATCAGATCCCGACACCCCAAAATGCTTCAGATGG 2485
1933 aLeuGlnGlnSerValGlyProThrValAsnGlnGlnProLeuGlnVal 1949
|||||
2486 TATCCAGGATCTGTAGCACACAGTGAATCAAGGAGCTTTGGAAAT 2535
1950 AlGlnValPheLeuAlaGlnLLeProGlnAspProLysLeuPheArgH 1966
|||||
2536 GCCAGGTTTTCGTGTGAATATACCTAGTACCAAGGCTCTTGACAGA 2585
1966 sHisAsnLysLeuArgLeuCysPheLysAspPheCysLysCysGln 1983
|||||
2586 TCATATTAATAACTGCACTGCTTTAAAGATTTACTTAAAGAGTGTGAAG 2635
1983 sPAlaLeuArgLysAsnLysAlaLeuLLeGlyProAspGlnLysGlnTyr 1999
|||||
2636 ATGCTTAAGAAAAATTAAGGCTTAATTTGGCGGTTCAAAAGGAGTAT 2685
2000 HisArgLuLeuGlnArgAsnTyrCysArgLeuArgGlnAlaLeuGln 2015
|||||
2686 CAAGGGAATGG...GGGAACATATCTTCGCTTAAGAGAGCCCTACAGC 2732
2016 ProLeuLeuThrGlnArgLeuProGlnLeuMetAlaProThr....Pr 2030
|||||
2733 CCT...AGATCACGAAGAGCCCTACGTTATCCAAAGCAATATGCTGT 2779
2030 oProGlyLeuArgAsnSerLeuAsnArgAlaSerPheArgLysAlaAsp 2047
|||||
2780 CCCCTGCCACAGAGATTTCTTCACTCAATGAGCTTTCCGAAAAATGATC 2829
2047 eu 2047
||
2830 TC 2831

```

seq\_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.ABL17737

seq\_documentation\_block:

ID ABL17737 standard; DNA; 5589 BP.

XX

AC ABL17737;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster genomic polynucleotide seq ID NO 4684.







```

929 GlnHisAlaIatPhePhePheGlnLeuMetValLysSerMetAlaLeuH1 945
      ::::::::::::::::::::|
2993 TCAAAATGGTGGTTTATTTAGCTATGTGCAAAATCATATGATGAAACA 3042
      |
945 sLeuLeuLeuGlnArgLeuAspThrProArgLysLeuArgPheProG 962
      |
3043 TTTCGATGTTCAAAATACTTTAAATGTCGCCGCAACATCATGATTTCCAC 3092
      |
962 LArgPheLeuAspAspIleThrAlaLeuValGlySerValGlyLeuGln 978
      |
3093 ATCAGTTTAATGATGATCTATCCACCCTGGTGCACCTAGTTACAAACAAA 3142
      |
979 ValIleThrArgValHisLysAspValGlyLeuAlaGlnHisLeuAsnAl 995
      |
3143 GGGGCGGCTACACACAGCAACCTTAATTAAGCGCATTTAAATATGC 3192
      |
995 aSerLeuAlaPhePheLeuSerAspLeuSerLeuValAspArgGlyP 1012
      |
3193 AAGTTTAGCCTTTTATTTGACATATCTAGCATCATGATGGGGAT 3242
      |
1012 heValPheSerLeuValArgAlaHisTyrLysGlnValAlaIleThrArgLeu 1028
      |
3243 TTGTTTGGCCTCATCAAAACGTACAAAGTTTGTGATTTCTAAAAAT 3292
      |
1029 GlnSerSerProAsnProAlaAlaLeuLeuThrLeuArgMetGluPheTh 1045
      |
3293 CCTCAATACCGGAT.....CTAATGAACATTAATAATAGACTTTT 3343
      |
1045 rArgIleLeuGlySerHisGlnHisTyrValIleThrLeuAsnLeuPro... 1060
      |
3343 GAGAAATGTGTGTAGTCAATGACATTTTGTAGCTTAAACTTGGCGTTTG 3383
      |
1061 .....CysCysProLeuSerProProAlaSerProSerProSerVal 1074
      |
3384 GTACTTCATACAAATGTAAACGCCATGCACGCCACGCCACGACACA 3433
      |
1075 sSerThrThrThrSerGlnSerThrPheSerSerGlnAlaProAspPr 1091
      |
3434 ACGTCGAGCAACAGCCAAACTTCTTGGGATCATAGAGAGA..... 3475
      |
1091 oLysValIleThrSerMetPheGlnLeuSerGlyProPheArgGlnGlnHisP 1108
      |
3476 .....GCCCTTCACGCCGCTTACTCAAGAAATCTCTCCACCAACAT 3518
      |
1108 heLeuAlaGlyLeuLeuLeuThrGluLeuAlaLeuAlaLeuGlu...Pro 1123
      |
3519 TTCTAGTCGGACTTGTCTGAGGACTTACAGAGTAATGAGAGTGCCA 3568
      |
1124 GluAlaGlnGlyAlaPheLeuLeuHisLysAlaIleSerAlaValAlaH1 1140
      |
3569 AACCCCA.....TTACATGGAAAGCAATTAAGGTATATCG 3606
      |
1140 sSerLeuLeuGlyHisAspThrAspProArgTyrAlaGluAlaIleThrV 1157
      |
3607 CAACCTGATGACATCTCAGATTTGGACGGCGGTATAGGAAACGATG 3656
      |
1157 aLysAlaArgValAlaGlnLeuTyrLeuProLeuLeuSerIleAlaArg 1173
      |
3657 CACGTGCAAGGGTGGCTGCTTATATATACCGCTACTTCTTAAGTGATG 3706
      |
1174 AspThrLeuProArgLeuHisAspPheAlaGlnGlyProGlnArgse 1190
      |
3707 GATTGATCTCTCATGTCATCGACAGCGCTTGACACAGATCGTGTGA 3756
      |
1190 rArgLeuAlaSerMetLeuAspSerAspThrGlnGlnGlyAspIleA 1207
      |
3757 GCAAAATCGGCACACTGAAGACATCAAGAGCGCACATCAACTATATACG 3806
      |
1207 IaGlyThrIleAsnProSerValAlaMetAlaIleAlaGlyLeuProLeu 1223
      |
3807 CATCAACCAATTAATTCGGAAGTGGCATTTGCATATCGGAAGTGTGCC 3856
      |
1224 AlaProGlySerArgAlaSerIleSerGlnGlyProProThrAlaSerAr 1240

```

```

3857 TACTCCTACTTGAATGACAGTAAACAAATTTGCC..... 3895
      |
1240 gAlaGlyCysAlaLeuSerAlaGlnSerSerArgThrLeuAlaCysV 1257
      |
3896 .....CTCAGCTCGGAGAAATACCCCATTTATTTGGTGTGTT 3932
      |
1257 aLeuThrValLeuLysAsnThrGluProAlaLeuLeuGlnArgThrAla 1273
      |
3933 TTCTGTGGGTCTAAATAAACTTGGACGCAACGTTTATATATGCTGGCTT 3982
      |
1274 ThrAspLeuThrLeuProGlnLeuGlyArgLeuLeuAspLeuTyrLe 1290
      |
3983 ATGATTTTAAGTCGCATAGAGTACGACCAAAATGCTTCAAGGTGTAACT 4032
      |
1290 uCysLeuAlaAlaPheGluTyrLysGlyLys.....AlaPheG 1304
      |
4033 TTGCTTGAATAAATTTGAGTACACAGCAACAAAAACGTGGCCACCTTA 4082
      |
1304 LuArgIleAsnSerLeuThrPheLysLys.....SerLeuAspMetLys 1318
      |
4083 AACGACGAAATACACAAAGTTTAGGAAACGTGATCTATGATGTAAAG 4132
      |
1319 AlaArgLeuGlnGluAlaIleLeuGlyThrIleGlyAlaArgGlnLume 1335
      |
4133 GAAAACTGGAAGAAATGTATTAGAGCACAAATTCAGCAGTATGATTT 4182
      |
1335 tValArgArgSerArgGluArgSerProPheGlnLysnProGluAsnVala 1352
      |
4183 AATAAATCTCGAAAGATCGTAAT.....TCAACAGAAAAATTC 4223
      |
1352 rGTPArgLys...SerValIleThrHisTyrLysGlnIleThrSerAspArgAl 1367
      |
4224 GATGAGGAAACACAGATCCATATCGCTCATATATGCGATCTGTG 4273
      |
1368 AspLysThrLysAspGlnMetGlnHisGlnAlaLeuValGlnGlyAsnLe 1384
      |
4274 GGAATAAGTACATGAACTGTAATTAAGTATCATTCATCGAAGCGCTT 4323
      |
1384 uAlaThrGluAlaSerLeuValValLeuAspThrLeuGlnIleLeuValG 1401
      |
4324 GCCAACTGAAGTTCGACTTGTCTGATACCTTGAATAATTTATGTTTC 4373
      |
1401 InThrValMetLeuSerGluAlaArgGlnSerValLeuGlyAlaValLeu 1417
      |
4374 ATGCTGCT.....GCCAACTTTATCAAAATCTTGGAACTGTGTA 4417
      |
1418 LysValValLeuTyrSerLeuGlySerAlaGlnSerAlaLeuPheLeuG 1434
      |
4418 AAGGTCTTCTTCATCTTGTCCCGGAATCAATCTGTTTGGCCCTGCA 4467
      |
1434 nHisGlyLeuAlaThrGlnArgAlaLeuValSerLysPheProGlnLeuL 1451
      |
4468 AAACCTGTTTGCCTCCCAACGGCGTTAATCTTTAAATTTCCAAACTTAT 4517
      |
1451 eupheGlnGluAspThrGlnLeuGlyCysAlaAspLeuGlyArgLeuLeu 1467
      |
4518 TATTGATGAGAGACGATATATGTGGGATTTGTGTCTAATTTTACTG 4567
      |
1468 ArgHisCysGlySerArgIleSerThrIleArgThrHisAlaSerAlase 1484
      |
4568 AAACATTTGTGGTCCCTATTGCCAGGAATACAGATCAGACAGCGCTGATC 4617
      |
1484 rLeuTyrLeuLeuMetArgGlnAsnPheGlnIleGlyHisAsnPheAla 1501
      |
4618 ACTATATTACTTAATGACAAAATTTTGAATTTGGAATTAAC..... 4660
      |
1501 rValLysMetGlnValThrMetSerLeuSerSerLeuValGlyThrThr 1517
      |
4660 ..... 4660
      |
1518 GlnAsnPheSerGlnGluHisLeuArgArgSerLeuLysThrIleLeuTh 1534

```

4660 ..... 4660  
1534 rTyrAlaGluGluAspMetGlyLeuArgSerThrPheAlaGluGlnV 1551  
4660 ..... 4660  
1551 aGlnAspLeuMetPheAsnLeuHisMetIleLeuThrAspThrValLys 1567  
4660 ..... 4660  
1568 MetLysGlnHisGlnGluAspProGluMetLeuIleAspLeuMetTyrAr 1584  
4660 ..... 4660  
1584 gTleAlaArgGlyTyrGlnGlySerProAspLeuArgLeuThrTrpLeug 1601  
4660 ..... 4660  
1601 LAsnMetAlaGlyLysHisAlaGluLeuGlyAsnHisAlaGluAlaAla 1617  
4660 ..... 4660  
1618 GlnCysMetValHisAlaAlaAlaLeuValAlaGluTyrLeuAlaLeuLe 1634  
4660 ..... 4660  
1634 uGluAspGlnArgHisLeuProValGlyCysValSerPheGlnAsnIleS 1651  
4660 ..... 4660  
1651 eRserAsnValLeuGluSerAlaIleSerAspAspIleLeuSerPro 1667  
4660 ..... 4660  
1668 AspGluGluGlyPheCysSerGlyLysHisPheThrGluLeuGlyLeuVa 1684  
4660 ..... 4660  
1684 IGLyLeuLeuGluGlnAlaAlaGlyTyrPheThrMetGlyGlyLeuTyrG 1701  
4660 ..... 4660  
1701 LuAlaValAsnGluValTyrLysAsnLeuIleProIleLeuGluAlaHis 1717  
4660 ..... 4660  
1718 ArgAspTyrLysLysLeuAlaAlaValHisGlyLysLeuGlnGluAlaPh 1734  
4660 ..... 4660  
1734 eThrLysIleMetHisGlnSerSerGlyTyrGluArgValPheGlyThrT 1751  
4660 ..... 4660  
1751 yRpheArgValGlyPheTyrGlyAlaHisPheGlyAspLeuAspGluGln 1767  
4660 ..... 4660  
1768 GluPheValTyrLysGluProSerIleThrLysLeuAlaGluIleSerHis 1784  
4660 ..... 4660  
1784 sArgLeuGluGluPheTyrThrGluArgPheGlyAspAspValAlaGluT 1801  
4661 ..... 4660  
1801 ..... TTTTACACTGACGATTCGGACCGACTCTGTGCATA 4697  
1801 leIleLysAspSerTyrProValAspLysSerLysLeuAspSerGlnLys 1817  
4698 TCATTAAAGATTCCAATACCGTTGATATTAATACCTTGATCCCGATAG 4747  
1818 AlaTyrIleGlnIleThrTyrValGluProTyrPheAspThrTyrGluLe 1834  
4748 GCTTACATTCATTAATGTGTAACCCCTACTTGAACATATGAAT 4797

1834 uLysAspArgValThrTyrPheAspArgAsnTyrGlyLeuArgThrPheL 1851  
1851 ..... 1851  
4798 GCGTCATCGTGACACATACCTTGACCGGAATTTCAATATAAACCGTTTA 4847  
1851 eurPheCysThrProPheThrProAspGlyArgAlaHisGlyGluLeuPro 1867  
4848 TATATGCCACGCCCTTTTACTAAATAAGGAAGCAGCATGGCGAATTAAT 4897  
1868 GlnGlnHisLysArgLysThrLeuLeuSerThrAspHisAlaPheProTy 1884  
4898 GAGCAGTGTAAACGGAAGAAAGATTTTGACACAGCGGATCACTTCCATA 4947  
1884 rIleLysThrArgIleArgValCysHisArgGluGluThrValLeuThrP 1901  
4948 CGTAAAAACCTGCATTTATGCTAATTAGCCAGACGCAATAGCTCGTGAAC 4997  
1901 roValGluValAlaIleGluAspMetGlnLysLysThrArgGluLeuAla 1917  
4998 CGATTGAAGTGGCAATTGAGATATTCAAAATAAATAACATTTGGAAATTGGCA 5047  
1918 PheAlaThrGluGlnAspProProAspAlaLysMetLeuGlnMetValLe 1934  
5048 GCTGCCACGATCAAGAGCCAGACAGCCCAATAATATTCGAATGTACT 5097  
1934 uGlnGlySerValGlyProThrValAsnGlnGlyProLeuGluValAlaG 1951  
5098 GCAGGATGTATTTGGAACCCATGTTAACCCAGACCGATGGAAATGGCGA 5147  
1951 InValPheLeuAlaGluIleProGluAspProLysLeu...PheArgHis 1966  
5148 GTGTGTCCTTTCCAAATTTATCCGACGGAACAACGTGTACCGCAAAACAC 5197  
1967 HisAsnLysLeuArgLeuGlyAspPheLysAspPheCysLysLysCysGluAs 1983  
5198 CAACAACAACCTCGGTGTGCTTGCAGAGTTTGCAGAGCTTGTCGCA 5247  
1983 rAlaLeuArgLysAsnLysAlaLeuIleGlyProAspGlnLysGlyTyrH 2000  
5248 TCGTTTGAATAAAGAAATCGCAATCTAATACTTTCAGATCAAAAAGATTAAC 5297  
2000 iAsArgLysLeuGluArgAsnTyrCysArgLeuArgGluAlaLeuGlnPro 2016  
5298 AACGAGAACTGGAACTATACACAGATCGGTTCATTTGAACGATTAAGTCCC 5347  
2017 LeuLeuThr 2019  
5348 TTTATTAAT 5356  
seq\_name: /SIDS1/gcdata/geneseq/geneseq-emb1/NA2001A.DAT:AAH43850  
seq\_documentation\_block:  
ID AAH43850 standard; cDNA: 2148 BP.  
XX  
AC AAH43850;  
XX  
DT 04-SEP-2001 (first entry)  
XX  
DE Preliminary human CLASP-7 encoding cDNA sequence.  
XX  
KW Human; CLASP-7; cadherin-like asymmetry protein; immune response;  
KW neuroprotective; antidiabetic; immunosuppressive; antirheumatic;  
KW antidiarrhetic; hypotensive; anti-HIV; cytostatic; immunostimulant;  
KW antihaemic; antiinflammatory; ophthalmological; neproretrophic;  
KW antihydroid; antiasthmatic; antiallergic; antibacterial; gene therapy;  
KW chromosome 19q13.2; autoimmune disease; multiple sclerosis; toxemia;  
KW juvenile diabetes; rheumatoid arthritis; pruritic urticarial papule;  
KW hypertension; Rh incompatibility; ss.  
OS Homo sapiens.  
XX  
Key Location/Qualifiers  
FH  
FT CDS 1..1932

```

FT      /*tag= a
FT      /partial
FT      /product= "preliminary CLASP-7 protein"
FT      /note= "no start codon given"
XX
XX      WO200142295-A2.
XX
XX      14-JUN-2001.
XX
XX      13-DEC-2000. 2000MO-US34152.
XX
XX      13-DEC-1999. 99US-0170453.
XX      14-JAN-2000. 2000US-0176195.
XX      14-FEB-2000. 2000US-0182296.
XX      11-APR-2000. 2000US-0196267.
XX      11-APR-2000. 2000US-0196460.
XX      11-APR-2000. 2000US-0196527.
XX      11-APR-2000. 2000US-0196528.
XX      11-APR-2000. 2000US-0547276.
XX      13-OCT-2000. 2000US-0240503.
XX      13-OCT-2000. 2000US-0240508.
XX      13-OCT-2000. 2000US-0240539.
XX      13-OCT-2000. 2000US-0240543.
XX
XX      (ARBO-) ARBOR VITA CORP.
XX
XX      Lu P, Garman JD, Candia AF;
XX      WPI: 2001-381641/40.
XX      P-PSDB; AAB99540.
XX
XX      Novel cadherin-like asymmetry protein-7 and polynucleotides encoding
XX      the polypeptide, useful for treating autoimmune disease,
XX      hypersensitivity, preventing transplant rejection by modulating immune
XX      response -
XX
XX      Example 3; Fig 1; 151pp: English.
XX
XX      The present invention describes a human cadherin-like asymmetry protein
XX      (CLASP), designated CLASP-7. The CLASP-7 protein (I) and its encoding
XX      nucleotide sequence (II) have activities including: neuroprotective;
XX      antidiabetic; immunosuppressive; antirheumatic; antiarthritic; anti-HIV;
XX      hypotensive; cytostatic; immunostimulant; antianemic; antiinflammatory;
XX      ophthalmological; nephrotoxic; antihypoid; antiaesthetic; anti-
XX      antiallergic; and antibacterial. (II) and CLASP-7 antibodies (III) are
XX      useful for detecting the CLASP-7 polypeptide. (II) is useful for
XX      producing (I) by recombinant methods. (I) or its fragment are useful for
XX      inhibiting an immune response in a cell such as T cell or B cell. A
XX      pharmaceutical composition (C), comprising (I) or (II), can be useful
XX      for treating CLASP-7-mediated disease such as an autoimmune disease
XX      caused or exacerbated by increased activity of T helper cells. Autoimmune
XX      diseases which can be treated using (C) include multiple sclerosis,
XX      juvenile diabetes and rheumatoid arthritis. (I) is useful for treating
XX      toxemia or pregnancy induced hypertension, pruritic urticarial papules
XX      and Rh incompatibility. (I) is also useful as a diagnostic reagent for
XX      immune and other disorders, since diseases characterized by
XX      overproduction or depletion of lymphocytes in blood or other organs may
XX      be detected by monitoring the level of (I) or its mRNA. CLASP-7 has been
XX      mapped to the chromosomal location 19q13.2. The present sequence
XX      represents the preliminary human CLASP-7 encoding cDNA sequence which is
XX      given in the present invention.
XX
XX      Sequence 2148 BP; 468 A; 649 C; 625 G; 406 T; 0 other.

```

## alignment\_scores:

```

Quality: 3326.00      Length: 643
Ratio: 5.181          Gaps: 0
Percent Similarity: 99.844      Percent Identity: 99.533

```

## alignment\_block:

```

US-09-736-968A-2 x AAH43850

```

```

Align seg 1/1 to: AAH43850 from: 1 to: 2148
1405 LeuSerGluAlaArgGluSerValLeuGlyAlaValLeuLysValValLeu 1421
|||||
1421 uTySerLeuGlySerAlaGlnSerAlaLeuPheLeuGlnHisGlyLeuA 1438
|||||
51 GTACAGCCTGGGAGAGTCCAGAGTCCCTCTTCTTGGAGCATGGCCCTGG 100
|||||
1438 laThrglAaArgAlaLeuValSerLysPheProGluLeuLeuPheGluGlu 1454
|||||
101 CCACCAGAGGAGCCCTTGTCTCAAGTTCAGAGTCTGTTCAGAGAG 150
|||||
1455 AspThrGluLeuCysAlaAspLeuLysLeuAlaGlyLeuLeuArgHisCysG 1471
|||||
151 GACAGGAGCTGTGTCCAGCTGTGCTGAGAGCTCTTACGACACTGTGG 200
|||||
1471 ySerArgIleSerThrIleArgThrHisAlaSerAlaSerLeuTyrrLeu 1488
|||||
201 CAGCCGCAATCAGACCATCCGACGACGACGACGACGCTGCTGACCTGC 250
|||||
1488 euMetArgGlnAsnPheGluIleGlyHisAsnPheAlaArgValLysMet 1504
|||||
251 TCATGCGACAGAACTTCGAGATCGGCGACAACTTGGCCGTGAAGATG 300
|||||
1505 GlnValThrMetSerLeuSerSerLeuValGlyThrThrGlnAsnPhase 1521
|||||
301 CAGGTCACCATGTCTCTCTGCTGCTGGGGGAGGAGCGACGACAACTTCAG 350
|||||
1521 rGluGlnHisLeuArgArgSerLeuLysThrIleLeuThrTyrrAlaGlu 1538
|||||
351 TGAAGAGCAGCTGCGAGCGTGTACCTCAAAACATCTCCACATATGCTGAG 400
|||||
1538 luAspMetGlyLeuArgAspSerThrPheAlaGlnValGlnAsnLeu 1554
|||||
401 AGCAGATGGGGCTGGCGGACGACACTTTCGCGACAGCTGCGAGAGCTGG 450
|||||
1555 MetPheAsnLeuHisMetIleLeuThrAspThrValLysMetLysGlnHis 1571
|||||
451 AGCTTCACCTGCACATGATCTGTGACGACGACGAGGAGGAGGAGGAGCA 500
|||||
1571 sGlnGluAspProGluMetLeuIleAspLeuMetTyrrArgIleAlaArg 1588
|||||
501 CCAGAGAGCCTGAGATGCTATCATGACCTCATGACGAATTCGCCGGG 550
|||||
1588 lYTyGlnGlySerProAspLeuAlaGlyLeuThrTrpLeuGlnAsnMetAla 1604
|||||
551 GCTACCGAGGGCTCACCGGACCTTCGCTGACCTGCTGCAGAACATGGCC 600
|||||
1605 GlyLysHisAlaGluLeuGlyAsnHisAlaGluAlaAlaGlnCysMetVal 1621
|||||
601 GGGAGGACGCGGAGACTGGGACACACCCAGCGCGCCAGTGCATGCT 650
|||||
1621 HisAlaAlaLeuValAlaGluTyrrLeuAlaLeuLeuGlnAspGlnA 1638
|||||
651 GCACCGCGCGCCCTCGCTGCTGAGTACCTGCGCTCGCAGAGCACACC 700
|||||
1638 rGHisLeuProValGlyCysValSerPheGlnAsnIleSerSerAsnVal 1654
|||||
701 GCCACCTGCGGCTGGGCTGCTTCTTCTTCAGAACATCTCATCCAAATGTG 750
|||||
1655 LeuGluGluSerAlaIleSerAspAspIleLeuSerProAspGluGlu 1671
|||||
751 CTAGAGGAGTCCGCCATCTCCGAGACATCTGTGCGCGGAGGAGGAGG 800
|||||
1671 yPheCysSerGlyLysHisPheThrGluLeuGlyLeuValGlyLeuLeuG 1688
|||||
801 CTTTGTCTCCGGGAAAGCACTTCAGTGTGGGCTGTAGGCTTCTGCG 850
|||||
1688 luGlnAlaIleGlyTyrrPheThrMetGlyGlyLeuTyrrGlyAlaValAsn 1704
|||||
851 AACAGGACGCGGCTACTTACCATGGGCGGCTCTTACGAGGCGGTGAAT 900
|||||

```

```
1705 GLuValTyrTysAsnLeuIleProIleLeuGluAlaHisArgAspTyrIy 1721
|||||
901 GAGGTCTACAAAGAAACCTCATCCCATCTCGAAGCCACCGTCACATCAA 950
|||||
1721 sLysLeuAlaIaValHisGlyTysLeuGlnGluAlaPheThrLysIleu 1738
|||||
951 GAAGCTGGCCGCGTGCACGGCAAACTGCAGGAGGCGCTTCACCAAGATCA 1000
|||||
1738 eThsGlnSerSerGlyTyrGluIaValPheGlyThrTyrPheArgVal 1754
|||||
1001 TGCACCGAGATCCGGCTGGAGGCGCTGTTCCGGACGTTTCCCGCGTG 1050
|||||
1755 GlyPheTyrGlyAlaHisPheGlyAspLeuAspGluGlnIuPheValTyr 1771
|||||
1051 GGCTTCTACAGCGCGCCACTTCGGTGACCTGGATGACGAGAGTTTGTTG 1100
|||||
1771 rLysGluPProSerIleThrTysLeuAlaGluIleSerHisArgLeuGlu 1788
|||||
1101 CMAAGAGCCATCGATCACAGAGCTGGCAGAGATCTCACACCGCTGGAG 1150
|||||
1788 IuPheTyrThrGluArgPheGlyAspAspValValGluIleIleLysAsp 1804
|||||
1151 AGTCTTACACGAGAGATTGGCCACACGCTGTTGAGATTATCAAGAC 1200
|||||
1805 SerTyrProValAspLysSerLysLeuAspSerGlnLysAlaTyrIleG 1821
|||||
1201 TCTTACCCTGTGACAAAGTCCAAAGCTTGACTCACAAAAGGCGTTACAT 1250
|||||
1821 nIleThrTyrValGluProTyrPheAspThrTyrGluLeuLysAspArg 1838
|||||
1251 GATCACGATGTGAAACCGTACTTGTGACTACAGAGCTCAAGACCGGG 1300
|||||
1838 aLThrTyrPheAspArgAsnTyrGlyLeuArgThrPheLeuPheCysThr 1854
|||||
1301 TGACCTACTTTGACCGCAACTATGGGCTTCGCACATTCCTGTCGACG 1350
|||||
1855 ProPheThrProAspGlyArgAlaHisGlyLeuProGluGlnHisLys 1871
|||||
1351 CCGTTCACGCGGATGGCGGCGCACACGGGAGCTGCCCGGCAACAA 1400
|||||
1871 sArgLysThrLeuLeuSerThrAspHisAlaPheProTyrIleLysThr 1888
|||||
1401 GCGTAGACGCGCTCTACACACCGACACGCGCTTCCCTACACTCAAGATC 1450
|||||
1888 rGileArgValCysHisArgGluGluThrValLeuThrProValGluVal 1904
|||||
1451 GCATCCGTGTGTGCCACCGGGAGAGAGCGTGTGACSCCAGTGGAGGTG 1500
|||||
1905 AlaIleGluAspMetGlnLysLysThrArgGluLeuAlaPheAlaThrG 1921
|||||
1501 GCCATCGAGGACATGCAGAAAGACACGAGGAGCTGGCGCTTGGCACCG 1550
|||||
1921 uGlnAspProProAspAlaLysMetLeuGlnMetValLeuGlnLysSer 1938
|||||
1551 GCAGAGACCCACCAATGCTTAAGATCTTACAGATGCTGCTTCAGGCGT 1600
|||||
1938 aGlyProThrValAsnGlnGlyProLeuGluValAlaGlnValPheLeu 1954
|||||
1601 TAGGGCCACCGTAAACCAAGGCTCCCTGGAGGTGGCCAGGTGTTTAA 1650
|||||
1955 AlaGluIleProGluAspProLysLeuPheArgHisAsnLysLeuArg 1971
|||||
1651 GCAGAGATCCCGAAGACCCCAAGCTCTTCGCGCATCAACAATAATTCG 1700
|||||
1971 gLeuCysPheLysAspPheCysLysLysCysGluAspAlaLeuArgLys 1988
|||||
1701 GCTTGTCTTCAAGGACTTCTGCAAGAATGTAGATGCGCTGCGGAGAAA 1750
|||||
1988 snLysAlaLeuIleGlyProAspGlnLysGluTyrHisArgGluLeuGlu 2004
|||||
1751 ATAAAGCCCTGATGTGGCGGACGAGAGAGAGTACCACCGTGTGCTGGAG 1800
```

```
2005 ArgAsnTyrCysArgLeuArgGluAlaLeuGlnProLeuLeuThrGlnArg 2021
|||||
1801 CGCAACTACTGCGCGCTGGGGAGGCTGTGCAGCCCGTGTACCCACG 1850
|||||
2021 gLeuProGlnLeuMetAlaProThrProProGlyLeuArgAsnSerLeuA 2038
|||||
1851 CCTGCCCGCAGCTGATGACACCCACCCGCGGCTTCAGGAATCTCTTGA 1900
|||||
2038 snArgAlaSerPheArgGlyAlaAspLeu 2047
|||||
1901 ACAGAGCAAGTTTCCGAAAGCGACACTC 1929
```

seq\_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA507382

seq\_documentation\_block:

ID AA507382 standard; cDNA: 4027 BP.

AC AA507382;

DT 26-SEP-2001 (first entry)

DE Human DNA associated with CLASP-5 #2.

KW Human; CLASP-5; cadherin-like asymmetry protein; immune gateway;

KW immunogen; antibody; autoimmune disease; rheumatoid arthritis;

KW multiple sclerosis; leukaemia; insulin dependent diabetes mellitus;

KW acquired immunodeficiency syndrome; AIDS; ss.

OS Homo sapiens.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

W0200142296-A2.

14-JUN-2001.

13-DEC-2000; 2000WO-US34163.

13-DEC-1999; 99US-0170453.

14-JAN-2000; 2000US-0176195.

14-FEB-2000; 2000US-0182296.

11-APR-2000; 2000US-0196267.

11-APR-2000; 2000US-0196460.

11-APR-2000; 2000US-0196527.

11-APR-2000; 2000US-0196528.

13-OCT-2000; 2000US-0240503.

13-OCT-2000; 2000US-0240508.

13-OCT-2000; 2000US-0240543.

13-OCT-2000; 2000US-0240539.

XX

XX

XX

XX

XX

(ARBO-) ARBOR VITA CORP.

Lu P, Garman JD, Candia AF;

WPI; 2001-367865/38.

P-PSDB; AAU04026.

XX

XX

XX

XX

XX

Disclosure: Fig 9C; 188pp; English.

The sequence encodes a protein presented as AAU04026, included in the specification which relates to a cadherin-like asymmetry protein, CC CLASP-5, which is a transmembrane protein of the immune system involved in the formation of the immune gateway. CLASP-5, polynucleotides encoding



1846	lyneuhgrhtrpheleuphecysthrprophethProAsp1yArGla	1862
1846		
2336	ACCTCGGAGGTTCAATGATACACCACCCCTTACCCCTGGAGGGGGGCGCT	2385
2336		
1863	HisGIgIuIeupProGIuInHisArglystrhrLeuSerThrAs	1879
1863		
2386	CGGGGAGAGCTGCATGACAGTAAAGAACACAGATCTTCACCATAT	2435
2386		
1879	pHisAlaPheProTyrIlelystrhrArgIleArgValCysHisArgGlu	1896
1879		
2436	GCAAGCCCTCCCTACATCAAGACCCGAGTACGCGTATCCGAAGAGAG	2485
2436		
1896	IuThrValIeupThrProValGIuValAlaIleGIuAspMetGIuInslys	1912
1896		
2486	AGTTTGTTTGACACCGATTGAAGTTGCCATTGAAGACATGAAGAAGAG	2535
2486		
1913	ThrArgGIuIeuaIaPheAlaThrGIuInAspProProAspAlaIysMe	1929
1913		
2536	ACCTCGAGTTAGCAGTGGCCATTACACAGAGCCCTCGATGCAAGAT	2585
2536		
1929	tLeuGIuMetValIeugIleGlySerValGIyProThrValAsnGIuIyP	1946
1929		
2586	GCTTCACATAGTGCTGCAGAGGCTCTGtGGAGCTACTGTAAATCAGGAC	2635
2586		
1946	roleuGIuValAlaGIuValPheLeuAlaGIuIleProGIuAspProIlys	1962
1946		
2636	CACGTGGAAGTAAAGCCCAAGTGTTTTGGCTGAATTCCTGTCATCCAAA	2685
2636		
1963	LeuPheArghHisHisAnIysLeuAlaGlyLeuCysPheIysAspPheCysIy	1979
1963		
2686	CTCTATGACATCCACAAACAAGTTGAGGTTATGCTTTAAGAAATTCATCAT	2735
2686		
1979	sIyCysGIuAspAlaLeuArGlyAsnIysAlaLeuIleGIyProAspG	1996
1979		
2736	GAGATGTGTGAGCTGTAGAGAAAAACAAGCCTTCATCAGCGCAGAC	2785
2736		
1996	IuIyGIuTyrHisArgGIuIeugIuIuArgAsnTyrCysArgLeuArGIu	2012
1996		
2786	AGAGGAATATCAGCAGAACTCAAAAAGAACTATAACAAGCTAAAGAG	2835
2786		
2013	AlaIeugIuProIeupThrGInArgLeuProGIuIeupMetAlaPro..	2028
2013		
2836	AACTCAAGCCCAATGATCGAGCGGAAAAATTCGAACTGTACAAAGCCAAAT	2885
2836		
2029	.....ThrProProGIuIeuaArgAsnSerLeuAsnArgAlaSerPhea	2043
2029		
2886	ATTCAAGATTGAGACTCAAAAGAGAGACTCTTCCACAGATCTAAGTTTCA	2935
2886		
2043	IgIyAlaIaAsp	2046
2043		
2936	GGAAATGTCAA	2946
seq_name:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA083334	
seq_documentation_block:		
ID	AA083334 standard; cDNA; 6454 BP.	
XX	AA083334:	
AC		
XX		
DT	23-OCT-2001 (first entry)	
XX		
DE	Human cDNA encoding CLASP-4.	
XX		
KW	Human: CLASP-4; cadherin-like asymmetry protein-4; autoimmune disease;	
KW	ataxia telangiectasia; Human immunodeficiency virus infection;	
KW	inflammatory disease; rheumatoid arthritis; multiple sclerosis;	
KW	diabetes mellitus; Immune disorder; Guillain Barre syndrome;	
KW	severe combined immunodeficiency; allergic reaction; asthma;	
KW	immunogen; antibody; ss.	
OS	Homo sapiens.	
XX		
XX		





[illegible]

```
644 .....ProArgProGlyThrAlaLeuGluThrProValGlyP 656
      :::::
2119 CACAAAGGAGACACCAAAAGACAGACAGCTGAACTCCAGTTGGGT 2168
      :::::
656 heThrTrpIleProLeuLeuGlnHisGlyArgLeuArgThrGlyProHe 672
      :::::
2169 TTGGCTGGGTACCTTTGCTGAAAAGATGAGAAATCATCATCATTTGAGCAG 2218
      :::::
673 CysLeuProValSerValAspGlnProPro..... 682
      :::::
2219 CAGCTGCGCAGTTTCCGCCAATCTTCCCGAGCTACTGATCTCAATGA 2268
      :::::
683 ...ProSerTyrSerValLeuThrProAspValAlaLeuProGlyMetA 698
      :::::
2269 TGCAGAAATCCAGAAAGCAATGTAACGTGATATT.....A 2303
      :::::
698 rgTrpValAspGlyHisIleGlyValAlaPheSerValGluLeuThrAlaVal 714
      :::::
2304 AATGGGTAGATGCTGCAAAAGCCTTTGTTGAAGTTTAAAGCCACTTAGAA 2353
      :::::
715 SerSerValHisProGlnAspProTyrLeuAspLysPhePheThrLeuVa 731
      :::::
2354 TCTACCATTTACACTCAAGATCTGCATGTCACAATCTTCCATCATG 2403
      :::::
731 HisValLeuGluGluGlyAlaPheProPheArgLeuLysAspThrValL 748
      :::::
2404 CCAGCTGATTCAGTCAAGCTCG..... 2425
      :::::
748 euSerGluGlyAsnValGluGlnLeuArgAlaSerLeuAlaLeu 764
      :::::
2426 .....AAGAAGTTCAGGAGGAGCTCATTAATATTAAAGTTTG 2467
      :::::
765 ArgLeuAlaSerProGluProLeuValAlaPheSerHisValLeuAs 781
      :::::
2468 CATGCCATGAGATCCCAAGTCATGATACAGTTCTACCTGTAATCTTAT 2517
      :::::
781 PheLeuValArgLeuValIleArgProProIleIleSerGlyGlnLeu 798
      :::::
2518 GCMACTCTCCGAGTTCACACAATATGACCCATGAAGATGACCTTCTTA 2567
      :::::
798 AlaLeuGlyArgGlyAlaPheGluAlaMetAlaHisValIleSerLeu 814
      :::::
2568 TCAC.....TGCACCATGGTCTCTTACATATTGATTCACAAAG 2605
      :::::
815 ValHisArgSerLeuGluAlaAlaGlnAspAlaArgGlyHisCysProG 831
      :::::
2606 TGCATGAA.....GAAGCTTGATAGTTA 2631
      :::::
831 nLeuAlaAlaTyrValHisTyrAlaPheArgLeuProGlyThrGluPro 848
      :::::
2632 TCTAAGATTCATCTCAAGATATAGCTTCGA.....CCTGAAAACCGA 2675
      :::::
848 eLeuProAspGlyAlaProProValThrValGlnAlaAlaThrLeuAla 864
      :::::
2676 GTGCTCTCTCAG..... 2686
      :::::
865 ArgGlySerGlyArgProAlaSerLeuTyrLeuAlaArgSerIle 881
      :::::
2686 ..... 2686
      :::::
881 eSerSerSerAsnProAspLeuAlaValAlaProGlySerValAspAsp 898
      :::::
2686 ..... 2686
      :::::
898 LuValSerArgIleLeuAlaSerLysLeuLeuHisGluLeuAlaLeu 914
      :::::
2687 .....GCCACCTGATACATGAACCCGCTACT 2716
      :::::
915 GlnTrp.....ValValSerSerSerAla.....ValArgGluAl 926
      :::::
2717 ACGATGATACCAATATTGAAACAGTCTGCAGATTTTTATCAATAAACAA 2766
      :::::

926 aLeLeuGlnHisAlaThrPhePheGlnLeuMetValLysSerMetA 943
      :::::
2767 ATTCTTAAGTACTCATGAGTTTCTTTGAAATATGCAAAATGCAATGG 2816
      :::::
943 LeuHisLeuLeuLeuGlyGlnArgLeuAspThrProGlyLysArg 959
      :::::
2817 CCACATAGTCTGTTGGAAGAGATTAAGATTAAAGTTCCCGAGGCCAGAGA 2866
      :::::
960 PheProGlyArgPheLeuAspAspIleThrAlaLeuValGlySerValG 976
      :::::
2867 TTTCCGAGACATATGATCATGTTTACATTCACCTGCTCTTCTTCAATAT 2916
      :::::
976 yLeuGluValIleThrArgValHisLysAspValGluLeuAlaGlnHis 993
      :::::
2917 TCCCATGATGACTATTCGTTATGCGAGATTCGCCGATGACTCAGAAATG 2966
      :::::
993 euAsnAlaSerLeuAlaPhePheLeuSerAspLeuSerLeuValAsp 1009
      :::::
2967 TGAACATTAAGTTGGCTAGACTCTCTGAAGCGCTGTTGACACTAATGAT 3016
      :::::
1010 ArgGlyPheValPheSerLeuValArgAlaHisTyrLysGlnValAlaTh 1026
      :::::
3017 AGAGGATTTATTTCAATTAATTAATGACTATATATGTGATTC..... 3061
      :::::
1026 rArgLeuGlnSerSerProAsnProAlaAlaLeuLeuThrLeuArgMetA 1043
      :::::
3062 .....AGCCCCAAGATCTTAAGTTCTGGCTGAATACAACTTG 3101
      :::::
1043 LuPheThrArgIleLeuCysSerHisGlnHisTyrValThrLeuAsnLeu 1059
      :::::
3102 AATTTGTGCAMACAAATTTCAATCAACGAAATTAATCATCTCTGAAACTTG 3151
      :::::
1060 ProCysCysProLeuSerProProAlaSerProSerProSerValSer 1076
      :::::
3152 CCAATG.....GCATTTGCAAAACCTTAATCTGACGG 3183
      :::::
1076 rThrThrSerGlnSerSerThrPheSerSerGlnAlaProAspProLys 1093
      :::::
3184 GGTTCAAATGTCAAATCTTGAATACGT..... 3211
      :::::
1093 aThrSerMetPheGluLeuSerGlyProPheArgGlnHisPheLeu 1109
      :::::
3212 .....TTATCAGATGAGTATTCGAAGCATCTCTTG 3244
      :::::
1110 AlaGlyLeuLeuLeuThrGluLeuAlaLeuAlaLeuGluProGluAlaG 1126
      :::::
3245 GTTGGTCTACTCTGAGGAACTTCATGCTCTTCCAGGACATTAATGA 3294
      :::::
1126 uGlyAlaPheLeuLeuHisLysAlaIleSerAlaValHisSerLeu 1143
      :::::
3295 G.....ATCAGATATACAGATCTCTGTATTAAGAATCTTT 3332
      :::::
1143 euCysGlyHisAspThrAspProArgTyrAlaGluAlaThrValLysAla 1159
      :::::
3333 TGATTAACATGATTTGACACAGATATACAGACACAGAACCAACACC 3382
      :::::
1160 ArgValAlaGluLeuTyrLeuProLeuLeuSerIleAlaArgAspThrLe 1176
      :::::
3383 AAAATAGCACAATGTACTCCCTTGTGTGACACTTTGAAAATAT 3432
      :::::
1176 uProArgLeu..... 1179
      :::::
3433 ACAGGATTAAGCAGTCCAGATACCTTGATCTTGTCAGCCATGCTTA 3482
      :::::
1180 .....HisAspPheAlaGluGly.....ProGlyGln 1188
      :::::
3483 ATTCTGCATCCAGAGATGAGTTTCCATGCTGGCTTACTTCACTGCCAAT 3532
      :::::
1189 ArgSerArgLeuAlaSerMetLeuAspSerAspThr..... 1200
      :::::
3533 AGAGGAACTGTGACACT.....GACAAAGACACCCCTTATGGGTCTT 3576
      :::::
1201 ...GluGlyGluGly.....AspIleAlaGlyThrIleAsn 1212
      :::::
```

```
3577 TCAAAATGGACATGGAATTAGAGAGAAGATTCAAGAGTTCCCTCATCC 3626
1212 roserValAlaMetAlaIleAlaGlyPro..... 1222
3627 CAGAGAGAGA.....ACAGATTTCCAGATCAGGCAACACTGGT 3667
1223 .....LeuAlaProGlySerAlaGlySerIleSerGlnGlyProPr 1236
3668 GAAATATACCCGACAGATCTTACAGAGAGTATCCAGATATAC... 3715
1236 oThrAlaSerArgAlaGlyCysAlaLeuSerAlaGluSerArgThrL 1253
3716 .....CGCTGGATCAGTATGAATCAGAGCC 3743
1253 euleuAlaCysValLeuThrPvalLeuLysAsn...ThrGluProAlaLeu 1268
3744 TCCGATGTCGTACCTGATATAGTAAATGATTCAGAGAGTACTGTC 3793
1269 LeuGlnArgThrAlaThrAspLeuThrLeuProGlnLeuGlyArgLeuL 1285
3794 TTAACTTACTGG...ATAAAGTATCACCCTCAGAGCTCATAAACATTC 3840
1285 uAspLeuLeuThrLeuCysLeuAlaAlaPheGluThrLysGlyLysAla 1302
3841 TATACTTTTGAAGTATGCTTTGTTCACTTTAGATATATGGGAAAGAA 3890
1302 laPheGluArgIleAsnSerLeuThrPheLysLys.....SerLeuAsp 1316
3891 ACATAGCAAGGTGCATGATGCTGGCTGCAAAACACTCGAATATAGAC 3940
1317 MetLysAlaArgLeuGlnGluAlaIleLeuGlyThrIleGlyAlaArgG 1333
3941 CGAAATCGCAAAACCATGCTGCTCTTCGAACAGATCAGAGATATGCA 3990
1333 ngLumetValArgArg...SerArgLuarSerProPheGlyAsnProG 1349
3991 GGGCCGGCTTCACATCTTAGTACCCAGAAAGTTTCATT..... 4030
1349 LysAsnValArgThrArgLysSerValThrHisTrpLysGlnThrSerAsp 1365
4031 .....ACACTTAATCACAGTTCTACAAACA..... 4054
1366 ArgValAspLysThrLysAspGluMetGlnHisGluAlaLeuValGlu 1382
4055 .....ACTGAAAGCAACATTTTCCACCAGGCACTTCTTGAAGG 4092
1382 yAsnLeuAlaThrGluAlaSerLeuValValLeuAspThrLeuGluIle 1399
4093 CAATATACGCTACGAGATTTCCTTAACAGTACAGACCATATCATTTT 4142
1399 leValGln.....ThValMetLeuSerGluAlaArgGluSer 1411
4143 TCACCTCAGTCTTCAAGACCAACTTTAAATATGATGGCCATAACCA 4192
1412 ValLeuGlyAlaValLeuLysValValLeuThrSerLeuGlySerAlaG 1428
4193 TTAAATGAAAAGGTGTTGATATACATCTTCTTTCTTAAATGAGACA 4242
1428 nSerAlaLeuPheLeuGlnHisGlyLeuAlaThrGlnArgAlaLeuValS 1445
4243 ATCTGAAGTGTGCTGAACATGATTTGGCTCAGCAGAGAGCTTTGATCA 4292
1445 erLysPheProGluLeuLeuPheGlnGluAspThrGluLeuCysAlaAsp 1461
4293 GTAAGTTTCTTCAGCATTTTCAAGAGAGATAAACATGTGTCTGCA 4342
1462 LeuCysLeuArgLeuLeuArgHisCysGlySerArgIleSerThrIleAr 1478
4343 TTTTGGATATAGGTTTAAAGTCTGCACATCGAAGATTTAGCTCAACAG 4392
1478 gThrHisAlaSerAlaSerLeuThrLeuLeuMetArgGlnAsnPhGlu 1495
1495 ..... 1495
```

---

```
4393 GATATAGACATCTGCACTTTGTATCTTTGATGAGAAACAACCTTGAGT 4442
1495 le.....GlyHisAsnPheAlaArgValLysMetGlnValThrMetSer 1509
4443 ATACCAAAAGAAAACCTTTTGAGAGACACATCTACAGATATATATGCT 4492
1510 LeuSerSerLeuValGlyThrThrGlnAsnPheSerGlnGlnHisLeuAr 1526
4493 GTAAACCAACGTATGCTGATGTACACTAAGCGAGATCAAAATTTTCA 4542
1526 garGSerLeuLysThrIleLeuThrTrpAlaGluGluAspMetGlyLeuA 1543
4543 GGAGCTTTATTCATTCATATCATATATTTTGCAATATGTACAGACCTATGA 4592
1543 rGAspSerThrPheAlaGluGlnValGlnAspLeuMetPheAsnLeuHis 1559
4593 AGCAACAGTCCCTTCCCGCAGAGTCAAAAGCTTGACCAAGAGATCCGC 4642
1560 MetIleLeuThrAspThrValLysMetLysGlnHisGlnGluAspProG 1576
4643 ACTGTTCTTATGGCCACCTGCCAAATGAAGAGATGAAGAAACCCCTGA 4692
1576 uMetLeuIleAspLeuMetLysArgIleAlaArgGlyThrGlnGlySerP 1593
4693 AATGCTAATGTATCTCCAGTATAGCTTAGCCAGTCCATGTCAAGACACC 4742
1593 roAspLeuArgLeuThrTrpLeuGlnAsnMetAlaGlyLysHisAlaGlu 1609
4743 CAGACCTCAGAAAACCTGGCTTGATATAGCATGGCCAAAGTATGATAAA 4792
1610 LeuGlyAsnHisAlaGluAlaAlaGlnCysMetValHisAlaAlaAlaLe 1626
4793 AATGAGATTTTTCAGAGGCTGCGATGTGTATGTCCATGTACGAGCTCT 4842
1626 uValAlaGluThrLeuAlaLeuLeuGlnAspGlnArgHisLeuProValG 1643
4843 AGTTCAGAGTTT.....CTTCATCGAAAATAATATTTCTCTAACG 4883
1643 LysCysValSerPheGlnAsnIleSerSerAsnValLeuGlnGluSerAla 1659
4884 GATGTTCCGCTTCAAGAAATTTACTCCCATATATGATGAAGAAAGACA 4933
1660 IleSerAspAspIleLeuSerProAspGluGlnGlyPheCysSerGlyL 1676
4934 ATGAAAGAGATGCTGGATGATGAT.....GT 4962
1676 shiSPheThrGlnLeuGlyLeuValGlyLeuLeuGlnGlnAlaAlaGlyT 1693
4963 CCATATATAGTGAAGAGTCTGCTGAGTTGCTAGAACAAATGTGTGATG 5012
1693 yRPheThrMetGlyGlyLeuThrGluAlaValAsnGluValThrLysAsn 1709
5013 GCTTTATGAGAGCAGAACGTTATGAATAATTTCTGAGATTTCCAAAGTTG 5062
1710 leuIleProIleLeuGluAlaHisArgAspTrpLysLysLysAlaAlaVala 1726
5063 ATCGTTCCAATTTATGAAACGTCGTGAGTTTGAGAACTTACTCAAGT 5112
1726 lHisGlyLysLeuGlnGlnAlaPheThrLysIleMetHisLysSerSerg 1743
5113 TTATATGACSTCTTATGAGAGCTTACACAAAATTTCTGAAATTTGCTATA 5162
1743 lYTPGluArgValPheGlyThrTrpPheArgValGlyPheTyGly... 1758
5163 CAAAAAGAGACTTTTAGGCACTTCTTCAAGAGTTGCTTTATAGGCCAA 5212
1759 AlaHisPheGlyAspLeuAspGlnGlnGluPheValThrLysGluProSe 1775
5213 TCTTTTGTGAAGAAAGATGGAAGAGATACATATTAAGAACCAAA 5262
1775 rIleThrLysLeuAlaGluIleSerHisArgLeuGlnGluPheThrG 1792
5263 GCTCACTGGCCCTCAGAAATTTCTGTAGACTTGTTAACCTTATATGCTG 5312
```





```

463 uAlaGluArgLeuSerAspGluAspLeuPheLysPheLeuAlaAspMetA 480
      ::::::::::::::::::::|
1701 CAGCAATTAAGCTATCCAAATGATGATGATGCTCAAGTACTTGCAAGCTTTC 1750
      ::::::::::::::::::::|
480 rGatArgProSerSerLeuLeuArgArgLeuArgProValThrAlaGlnLeu 496
      ::::::::::::::::::::|
1751 GGAACACCTGGAGAG...ATGGCTAAGCTCCCACTGATTTTGAAGCAATCTA 1797
      ::::::::::::::::::::|
497 LysLeuAspLeu...SerProAlaProGluAsnProHisPheCysLeuSe 512
      ::::::::::::::::::::|
1798 GACATTTCATTTGATATATGTTTCTCCAGACTTCCCTTAATTAATGTTAATTC 1847
      ::::::::::::::::::::|
512 rProGluLeuLeuHisLLeuLysProTyrProAspProArgGlyArgProT 529
      ::::::::::::::::::::|
1848 ATCATACATT.....CCCA 1861
      ::::::::::::::::::::|
529 hrlYsgLile.....LeuGluPheProAlaArg 538
      ::::::::::::::::::::|
1862 CAAACACATTGAAACCTGCAGTAAACCTCCCATCACGTTTGAAGTGGAG 1911
      ::::::::::::::::::::|
539 GluValTyr.....AlaProHisThrSerTyrAr 548
      ::::::::::::::::::::|
1912 GAATTGTGCGCTGCATACCAAAACACACCTCACGCTTACACCATTCACAC 1961
      ::::::::::::::::::::|
548 gaSnLeuLeuTyrValTyrProHisSerLeuAsnPheserSerArgGln. 564
      ::::::::::::::::::::|
1962 CAATCACCTTTACGTTTATCCCTAAGTACTTGAATATGACAGCTCAGAGT 2011
      ::::::::::::::::::::|
565 ....GlySerValArgAsnLeuAlaValArgValGlnTyrMet...Thr 578
      ::::::::::::::::::::|
2012 CTTTTCGCAAGGCTAGAAATATTCGATTGCAATTCAATTCAAAGATTCA 2061
      ::::::::::::::::::::|
579 GlyGluAspProSerGlnAlaLeuProValLLePheGlyLysSerSerCy 595
      ::::::::::::::::::::|
2062 GATGAGGAGAACCTCGAGCCCTTAAGTCATTATGCGCAGACCTGCTGG 2111
      ::::::::::::::::::::|
595 sSerGluPheThrArgGluAlaPheThrProValValTyrHisAsnLys 612
      ::::::::::::::::::::|
2112 GCCAGTTTTCACAGAGAGCGCTTGTGCGATTTTACACACATCACCAAA 2161
      ::::::::::::::::::::|
612 ePProGluPheTyrGluGluPheLysLeuHisLeuProAlaCysValAlThr 628
      ::::::::::::::::::::|
2162 ACCCAGAAATTATGATGAGATTAATAATAGAGTGGCCCATCGCTCGCAT. 2211
      ::::::::::::::::::::|
629 GluAsnHisLLeuLeuPheThrPheTyrHisValSerCys..... 642
      ::::::::::::::::::::|
2212 GAAACACACACCTGTGCTCACATTTCTCCATGTCAGCTGGACAACTC 2261
      ::::::::::::::::::::|
643 .....GlnProArgProGlyThrAlaLeuGlnThrProValGlyP 656
      ::::::::::::::::::::|
2262 AAGTAAAGCAGACGAAAGAGGATGCTGTTGAACCCAAAGTTGGCT 2311
      ::::::::::::::::::::|
656 heThrTyrLLeuPheLeuGlnHisGlyArgLeuArgThrGlyProPhe 672
      ::::::::::::::::::::|
2312 ACTCCGTGGCTCCCTCGTGAAGAGCGAAGGTTGGTCAAGCGCAG 2361
      ::::::::::::::::::::|
673 CysLeuProValSerValAspGlnProProProSerTyrSerValLeuTh 689
      ::::::::::::::::::::|
2362 CACATCCCGTCTCGGGAACCTT.....CCTTCGGGCTATCTTGCTCA 2405
      ::::::::::::::::::::|
689 rProAspValAlaLeu.....ProGluMetArgTyrValA 701
      ::::::::::::::::::::|
2406 CCGAGGAGCTTGGATGGCAGGCAATATGCTCCGGAATTAATTAATGGGATG 2455
      ::::::::::::::::::::|
701 spGlyHisLysGlyValPheSerValGluLeuThrAlaValSerSerVal 717
      ::::::::::::::::::::|
2456 ATGGAGCAGCAAGCACTGCGAAATTTCCATCATCTGTTTCTTCACAG 2505
      ::::::::::::::::::::|
718 HisProGlnAspProTyrLeuAspLysPhePheThrLeuValHisValLe 734
      ::::::::::::::::::::|
2506 TATACTCAGGATCAGCATTTACATTAATTTTTCAGTACTGTCAGAAAAAC 2555
      ::::::::::::::::::::|
734 uGluGluGlyAlaPheProPheArgLeuLysAspThrValLeuSerGlu 751

```

```

2556 CCAATCTGAGACC.....CAAGCTTAG 2578
      ::::::::::::::::::::|
751 LysAsnValGluGlnLeuArgAlaSerLeuAlaAlaLeuArgLeuAla 767
      ::::::::::::::::::::|
2579 GAAC.....GACTTGTAAAGTACCTTAAGAGCTCATGCGCATG 2619
      ::::::::::::::::::::|
768 SerProGluProLeuValAlaPheSerHisHisValLeuAspLysLeuVal 784
      ::::::::::::::::::::|
2620 GAAGCCACAGTATGATCGCTTCTTGCCCACTATCCAAACAGCGTGT 2669
      ::::::::::::::::::::|
784 LArgLeuValLLeuArgProProLLeSerGlyGlnLLeValAsnLeuG 801
      ::::::::::::::::::::|
2670 CCGAGTCTCCACACAGA...GCCACACAGAGAAGATCGCGTTAAGCTGA 2716
      ::::::::::::::::::::|
801 LArgGlyAlaPheGluAlaMetAlaHisValValSerLeuValHis... 816
      ::::::::::::::::::::|
2717 CTCGG.....GTCAATTATTCATGTGTGTTGCCAGCTGCCATGAG 2754
      ::::::::::::::::::::|
817 ArgSerLeuGluAlaAlaGlnAspAlaArgGlyHisCysProGlnLeuAl 833
      ::::::::::::::::::::|
2755 GAAGATTTGGAGAC.....CACTTGAG 2777
      ::::::::::::::::::::|
833 aAlaTyrValHisTyrAlaPheArgLeuProGlyThrGluProSerLeuP 850
      ::::::::::::::::::::|
2778 GTCAATATGTTAAGTACGCGTATAG.....GCTGAGCCATATGTT. 2817
      ::::::::::::::::::::|
850 roAspGlyAlaProProValThrValGlnAlaAlaThrLeuAlaArgGly 866
      ::::::::::::::::::::|
2817 ..... 2817
      ::::::::::::::::::::|
867 SerGlyArgProAlaSerLeuTyrLeuAlaArgSerLysSerLeSer 883
      ::::::::::::::::::::|
2818 .....GCCTCGAATAC..... 2829
      ::::::::::::::::::::|
883 rSerAsnProAspLeuAlaValAlaProGlySerValAspAspGluValS 900
      ::::::::::::::::::::|
2829 ..... 2829
      ::::::::::::::::::::|
900 eArgGlyLeuAlaSerLysLeuHisGluGluLeu.....AlaLeu 914
      ::::::::::::::::::::|
2830 .....AAGACATGCATGAAGAACTGACCAATATCATG 2862
      ::::::::::::::::::::|
915 GlnTyrValValSerSerAla.....ValArgGluAlaLLele 928
      ::::::::::::::::::::|
2863 ACCACGATTTCAAGCCCTTGTCCGATTTCTCTCCACGCAACAAACTACT 2912
      ::::::::::::::::::::|
928 uGlnHisAlaTrrPhePhePheGlnLeuMetValLysSerMetAlaLeu 945
      ::::::::::::::::::::|
2913 GAAGTACATGATGTTTCTTGTGATGTAAGTCAATCTATGCTCAGC 2962
      ::::::::::::::::::::|
945 LLeuLeuLeuGlyGlnArgLeuAspThrProArgLysLeuArgPhePro 961
      ::::::::::::::::::::|
2963 ATTGATATGAGAACTCCAAAGTAAATGCTGCGCAACCAAGATTTCT 3012
      ::::::::::::::::::::|
962 GlyArgPheLeuAspAspLLeThrAlaLeuValGlySerValGlyLeuG 978
      ::::::::::::::::::::|
3013 GCATCCTATCATTCATGCAAGTGAACCGTTGTAAATATGCTGATGCCACA 3062
      ::::::::::::::::::::|
978 uValLLeThrArgValHisLysAspValGluLeuAlaGlnHisLeuAsn 995
      ::::::::::::::::::::|
3063 CATCACTCAGAAAGTTTGAGATTAATCCAGAGGCATCTAAGAACGCGATC 3112
      ::::::::::::::::::::|
995 LAsrLeuAlaPhePheLeuSerAspLeuSerLeuValAspArgGly 1011
      ::::::::::::::::::::|
3113 ATACGCTTGGCTGTTCATCAAGAGATGTTTCACTTACATGACAGAGGCG 3162
      ::::::::::::::::::::|
1012 PheValPheSerLeuValArgAlaHisTyrLysGlnValAlaThrArgLe 1028
      ::::::::::::::::::::|
3163 TTGTCTTTC.....AAGCAGATCAACAAACTGCAT 3191
      ::::::::::::::::::::|
1028 uGln.....SerSerProAsnProAlaAlaLeuLeuThrLeuArgMetS 1043
      ::::::::::::::::::::|

```



seq\_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/NA2000.DAT.AAC74524

seq\_documentation\_block:

ID AAC74524 standard; cDNA; 6816 BP.

XX AAC74524;

XX

XX 08-FEB-2001 (first entry)

XX

DE Human ORF79 polynucleotide sequence SEQ ID NO:157.

XX

XX Human; open reading frame; ORF; detection: cytosatic; hepatotropic;

XX vulnereary; antisporadic; antiparkinsonian; nootropic; neuroprotective;

XX anticonvulsant; osteopthic; antilarthritic; immunosuppressant; cardiant;

XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;

XX hypotensive; dermatological; immunosuppressive; antinflammatory;

XX antiviral; antibacterial; antifungal; antirheumatic; antihypertoid;

XX antianemic; gene therapy; cancer; proliferative disorder; hypertension;

XX neurodegenerative disorder; osteoarthritis; graft vs host disease;

XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;

XX cholesterol ester storage; systemic lupus erythematosus; infection;

XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;

XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

XX bone damage; cartilage damage; antinflammatory disease; coagulation;

XX thrombosis; contraceptive; ss.

XX

5720 TCCGCGCTTCATGTTGAGATGSCATTATACGACACCGGAGAGACG 5769

1864 GlytLeuLeuProGluGlnHisLysArgLysThrLeuLeuSerThrAsp1 1880

5770 GCGCGGGGGAAGAGCAGTCGCAACGCGCGCACCATCTCGACAGCATACA 5819

1880 sAlaPheProTyrIleLysThrArgIleLeuValCysHisArgGluGly 1897

5820 CTGCTTCCTCATATGTGAAGAAGCCATCCCTGTATATACAGACACACA 5869

1897 hrValLeuThrProValGluValAlaIleGluLysPheGlnLysThr 1913

5870 CTGACCTGAACCCCATCGAGTGGCCATGTACGAGATGAGTAAGAAGGT 5919

1914 ArgIleLeuAlaPheAlaThrGluGlnAspProProAspAlaLysMet 1930

5920 GCGGAGGCTCCGGCAGCTGTCTCTCCGGCAGAGTGACATGTCAAACT 5969

1930 uGlnMetValLeuGlnGlnLysSerValGlyProThrValAsnGlnGlyPro 1947

5970 GCACCTCAAACTCCAGGGCAGCGGTGATCTTGACGTCAATGCTGGCCAC 6019

1947 eugIuValAlaGlnValPheLeuAlaGluIleProGluLysProLysLeu 1963

6020 TAGCATATGCGCGAGCTTCTTA.....GATGATCAACACACA 6057

1964 PheArgHis...HisAsnLysLeuArgLeu.....CysPheLysAs 1976

6058 AAGCGATATCTCTGACATTAAGTAAGAGCTCTTAAGGAAGTTTCAGGCA 6107

1976 pPheCysAlaLysCysGlnLysAlaLeuArgLysAsnLysAlaLeuIleG 1993

6108 ATTGTGTGAGAGCTTGGCGTCAACCTTAGCGGTAAACGAACGTCGATTA 6157

1993 LyrProAspGlnLysGluTyrHisArgGluLeuLysArgAsnTyrCysArg 2009

6158 AAGAAGACCCAGCTCGAGATATCAAGAAAGAAATAAACCCATCAAGGAA 6207

2010 LeuAlaArgIuAlaLeuGlnProLeuLeuThrGlnArgLeuProGlnLeu 2026

6208 ATGGCGAGAGAGCTTCTGTGAATCATGACATGACACATCTGCCCCCTGGA 6257

2026 tAlaPThrProProGlyLeuAlaArgAsnSerLeuAsn 2038

6258 GGAGAAAGACGAGCGTC...TTACCGAATTCCTCCCTTAC 6291









```
3973 CAGCAGTCCCTGCATCATCAACACGCTGCACACAGTGCACCGCTTAT 4022
1542 uArgAspSerThrPheAlaGluGlnValGlnAspLeuMetPheAsnLeuH 1559
4023 TAAGCACACCAAGCTTCTCTGATGTGAAGACCTTAAACCAAAAGGATAC 4072
1559 lMetIleLeuThrAspThrValLysMetLysGlnHisGlnGluAspPro 1575
4073 GCACGGTGTATAGGCCACCGCCAGATGAAGAGCATGAGACACGCCA 4122
1576 GluMetLeuIleAspLeuMetLysArgLLeuAlaArgGlyTyrGlnGlySe 1592
4123 GAGATGCTGTGAGACCTTCACAGCTGCACAAATCTATGCCACAC 4172
1592 rProAspLeuArgLeuThrTyrPheGlnAspMetAlaGlyLysHisAlaG 1609
4173 GCCCAGCTCAGGAAGACGTGCTGCACAGCATGGCCAGATCCATGTCA 4222
1609 luleuGlyAsnHisAlaGlnAlaAlaGlnCysMetValHisAlaAlaAla 1625
4223 AAAATGGCGATCTCTCAGAGCAGCAATGTCTATGTCCAGTAAACAGCC 4272
1626 leuValAlaGluTyrLeuAlaLeuGlnAspGlnArgHisLeuPro.. 1641
4273 CTAGTGGCAATATCTCACAGGAAGAACGATCCAGTGGAGCCGCC 4322
1641 ..... 1641
4323 CCTTCTCCCCACAGCCATAGCGCTGCTGAGAGAGACCGGGAGCGC 4372
1642 ..... ValGlyCysValSerPheGlnAsnIleSerSerAsnValLeu 1655
4373 TGTTTAGCAAGAGATGCACCGCTTCAAGGGCTATACCCCAACATCGAC 4422
1656 GluGlnSerAlaIleSerAspAspIleLeuSerProAspGlnGluGlyPh 1672
4423 GAGAGGCTCTCATGATGGAAGACGTGGGATGCAGAT..... 4461
1672 ecysSerGlyLysHisPheThrGlnLeuGlyLeuValGlyLeuGlnGlu 1689
4462 ..... GTCCATTTCAAGAGAGATGTCCTGATGAGGCTCTTAGC 4501
1689 lnaIlaAlaGlyTyrPheThrMetGlyLeuTyrGlnAlaValAsnGlu 1705
4502 AGTCGGCAGATGACCTCGAAGACCGGCTTACAGCTCATTCATGCCGAC 4551
1706 ValTyrLysAsnLeuIleProIleLeuGlnAlaHisArgAspTyrLysLy 1722
4552 ATCTACAAACTTATCATCCCATTTATGAGAACGGAGGATTTTGAGAG 4601
1722 sleuAlaAlaValHisGlyLysLeuGlnGluAlaPheThrLysIleMetH 1739
4602 GCTGGCCCATCTGTATGACACGCTGCACCGGCTTACAGCAAGTGCACG 4651
1739 lsgInserSerGlyTyrPgluArgValAlaPheGlyThrTyrPheArgValGly 1755
4652 AGGTCATGCACTCGGCGCGAGGCTTCTGGGAGCTACTTCGGGTGCC 4701
1756 pheTyrGly..... 1758
4702 TTTCTTGGGAGGAGCGCATACCACTTAAACAGACAGTGAACAGATGT 4751
1759 ... AlahisPheGlyAspLeuAspGlnGlnGluPheValTyrLysGluP 1774
4752 GGAGGGATTTCTTGAAGATGAAGATGAAGAGATATATTACAAAGAAC 4801
1774 rosetIleThrLysLeuAlaGluIleSerHisArgLeuGlnGluPheTyr 1790
4802 CCAATCTACACCGCTGTGGAAATTTCTCAGAGACTCTTAAGACTGTAC 4851
1791 ThrGluArgPheGlyAspAspValValGluIleLeuLysAspSerTyrPr 1807
4852 TCGGATAAATTGGTCTGAAAAATGTCAAAATGATACAGATTCGGCAA 4901
1807 oValAspLysSerLysLeuAspSerGlnLysAlaTyrIleGlnIleThrT 1824
4902 GGTCAACCTTAAGAGATGTGATTTTAAGTATGCTTACATCCAGGTACTC 4951
1824 yValGluProTyrPheAspThrTyrGlnLeuLysAspArgValThrTyr 1840
4952 ACSTCATCCCTCTTGTGAGAAAAAGAGTTGCAAGAAAGAAACAGAG 5001
1841 PheAspArgAsnTyrGlyLeuArgThrPheLeuPheCysThrProPheH 1857
5002 TTTGAGAGATCCCAACATCCGCCCTTCAATGTTGAGATCCATTTAC 5051
1857 rProAspGlyArgAlaHisGlyLeuProGlnGlnHisLysArgLysT 1874
5052 GCAGACCGGGAAAGACGACGGCGGGTGAAGAGCATGCCAAACGGCGA 5101
1874 hrLeuLeuSerThrAspHisAlaPheProTyrIleLysThrArgIleArg 1890
5102 CCATCCTGACAGCATACACTGCTTCCCTTATGTGAAGAACGCCATCCCT 5151
1891 ValCysHisArgGlnGluThrValLeuThrProValGluValAlaIleG 1907
5152 GTCATGTACCAACACCACTGACCTGAACCCCATCGAGTGGCCATTTGA 5201
1907 uAspMetGlnLysLysThrArgGluLeuValAlaPheAlaThrGlnGlnAsp 1924
5202 CGAGATGAGTAAAGAGTGGCGGAGCTCCGGACAGCTGCTCTCGGCCG 5251
1924 roProAspAlaLysMetLeuGlnMetValLeuGlnGlySerValGlyPro 1940
5252 AGGTGACATGATCAACACTGACCTCAACCTCCAGGCGAGCTGAGTGT 5301
1941 ThrValAsnGlnGlyProLeuGlnValAlaGlnValPheAlaGluIle 1957
5302 CAGGTAAATGCTGGCCCTACTAGCATATGCGCGAGCTTCTTA..... 5343
1957 eProGluAspProLysLeuPheArgHis...HisAsnLysLeuArgLeu. 1972
5344 ... GATGATACAAACACAAAGCATATCTTGACATAAAGTGAAGCTGC 5389
1973 ..... CysPheLysAspPheCysLysLysCysGlnAspAlaLeuArg 1986
5390 TTAAGGAACTTTTACAGCAATTTGTGAAGCTTGGCGTCAACCTTACCG 5439
5440 GTPAAAGAACGCTGTGATTTAAAGAACGACGCTCGAGTATCAGAGAAAT 5489
2003 uGluArgAsnTyrCysArgLeuArgGlnAlaLeuGlnProLeuLeuThrG 2020
5490 GAAAGCAACTACAGGAAATGCGCAAGAGACTTCTGAAATCATGCAATG 5539
2020 lArgLeuProGlnLeuMetAlaProThrProPogLysLeuArgAsnSer 2036
5540 AGGAGATCTGGCCCCCTGGAGGAAAGACGAGCTC...TTACGAAATTC 5586
2037 leuAsn 2038
5587 CTTCAC 5592
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH95572
seq_documentation_block:
ID AAH95572 standard; cdna; 1652 bp.
XX
AC AAH95572;
XX
DT 16-Oct-2001 (first entry)
XX
DE Human protein encoding cdna sequence seq ID NO:407.
XX
```

Human: cancer; ulcer; HIV infection; human immunodeficiency virus; antiinflammatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianemic; anaemia; antiaggregant; haemostatic; vulnery; antilucer; osteopathic; eczema; dermatological; antiallergic; antisthmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; infection; immunostimulant; gene therapy; antisense therapy; vaccine; inflammation; antiaphyllactic; rheumatoid arthritis; septic shock; pancreatitis; cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmune; genetic disease; haematopoietic disorder; platelet disorder; asthma; thrombocytopenia; osteoporosis; severe combined immunodeficiency; allergic rhinitis; diabetes; multiple sclerosis; depression; Alzheimer's disease; Parkinson's disease; neurodegenerative disorder; neurological disorder; ss.

Alzheimer's disease; Parkinson's disease; neurodegenerative disorder; neurological disorder; ss.  
Homo sapiens.  
MO200153455-A2.  
26-JUL-2001.  
22-DEC-2000; 2000WC-US35017.  
23-DEC-1999; 99US-0471275.  
21-JAN-2000; 2000US-0488725.  
25-APR-2000; 2000US-0552317.  
(HXSE-) HXSEQ INC.  
Tang YT, Liu C, Dymnac RT;  
WPI: 2001-457603/43.  
P-PSDB; AAM25631.  
Isolated human polynucleotides encoding polypeptides, useful for the treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
Claim 1; Page 505; 1217pp; English.

AAH99166 to AAH99934 encode the human proteins given in AAM25225 to AAM25963. The proteins can have activities based on the tissues and cells they are expressed in, such as: antiinflammatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianemic; antiaggregant; haemostatic; vulnery; antilucer; osteopathic; dermatological; antiallergic; antisthmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; and immunostimulant. The proteins and polynucleotides encoding them can be used in gene therapy, antisense therapy and vaccine production. The proteins and polynucleotides are useful for screening for agonists or antagonists of a protein and for the treatment and diagnosis of disorders associated with the activity of a protein e.g. inflammation, rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction, neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal infections, autoimmunity, genetic diseases, haematopoietic disorders, anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers, osteoporosis, severe combined immunodeficiency, eczema, allergic rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression, Alzheimer's disease, Parkinson's disease, neurodegenerative and neurological disorders.

Sequence 1652 BP; 377 A; 478 C; 479 G; 318 T; 0 other;

alignment\_scores:  
quality: 2195.00 length: 427  
ratio: 5.177 gaps: 2  
Percent Similarity: 99.297 Percent Identity: 99.063

alignment\_block:  
US-09-736-968a-2 x AAH99572 ..

Align seg 1/1 to: AAH99572 from: 1 to: 1652

1623 AAlaAlaLeuValAlaGluTyrLeuAlaLeuGluLysGlnArgH 1639  
2 GCGGCGCGCTCGTGGCTGAGTACCTCGCCCTGCTCAGAGCACCGCCA 51  
1639 sLeuProValGlyCysValSerPheGlnAsnIleSerSerAsnValLeuG 1656  
52 CCGGCCCGTGGCGCTGCTTCAGAAATCTCATCAATGCTGCTAG 101  
1656 lGluSerAlaIleSerAspAlleLeuSerProAspGluGluGlyPhe 1672  
102 AGGAGTCCGCGCATCTCCGACGACATCTCTGCGCCGACGAGAGGCTTC 151  
1673 CysSerGlyLysHisPheThrGluLeuGlyLeuValGlyLeuGluGlu 1689  
152 TGCTCCGGGAGACACTTCATGAGCTGGGCGCTGTAGGGTGTGTGAGACA 201  
1689 nAlaAlaGlyTyrPheThrMetGlyGlyLeuTyrGluAlaValAsnGlu 1706  
202 GGCAGCCGCGTACTTCACATGGCGGCGCTCTACAGGCGGTGATAGG 251  
1706 aTyrLysAsnLeuIleProIleLeuGluAlaHisArgAspTyrLysLys 1722  
252 TCTACAAAGAACCTCATCCCATCTCGAAGCCACCGCTGACTACAAAGAG 301  
1723 LeuAlaAlaValHisGlyLysLeuGlnGluAlaPheThrLysIleMetH 1739  
302 CTGGCCGCGGTGCACAGCAAACTGCAGAGCGCTTCACCAAGATCATGCA 351  
1739 sGlnSerSerGlyTyrProGluArgValPheGlyTyrThrPheArgValGly 1756  
352 CGAGGTTCGCGCTGGAGCGGCTGTGGAGCTATTTCCGCGTGGCT 401  
1756 heTyrGlyAlaHisPheGlyAspLeuAspGluGlnGluPheValTyrLys 1772  
402 TCTACGCGCGCCACTGTGGTGACTCGATGAGCAGAGATTGTGTACAG 451  
1773 GluProSerIleThrLysLeuAlaGluIleSerHisArgLeuGluGluP 1789  
452 GAGCATGATGATCAGCAACCTGGCAGAGATCTCACACCGCGTGGAGGATT 501  
1789 eTyrThrGluArgPheGlyAspAspValValGluIleIleLysAspSer 1806  
502 CTACACGAGAGATTGTGGCGAGCGTGTGAGATTAATCAAACTCTA 551  
1806 yProValAspLysSerLysLeuAspSerGlnLysAlaTyrIleGlnIle 1822  
552 ACCGTGTGACAAAGTCCAGCTTGACTCACAAMAGGCTCATCAGATC 601  
1823 ThrTyrValGluProTyrPheAspThrTyrGluLeuLysAspArgVal 1839  
602 ACGATGTGGAACCTGATTTGATTACTACAGCTCAAGGACCGGGTAC 651  
1839 rTyrPheAspArgAsnTyrGlyLeuArgThrPheLeuPheCysThrPro 1856  
652 CTACTTTAGCCGCAACTATGSGCTTCGCACATTTCTTTCGACGCGCT 701  
1856 heThrProAspGlyArgAlaHisGlyLeuLeuProGluGlnHisLysArg 1872  
702 TCACGCCGATGGCGCGCACACGCGGAGCTGCCGAGCAACACACGCT 751  
1873 LysThrIleLeuSerThrAspHisAlaPheProTyrIleLysThrArgI 1889  
752 AAGAGCTGCTCAGACCGACACGCTTCCCTCACTCAAGAGCTCCAT 801  
1889 eArgValCysHisArgGluGluThrValLeuThr.ProValGluValAla 1905  
802 CCGGTGTGCGCACCGGAGGAGAGGTGTGTACGCCCAAGTGGAGTGGCC 851  
1906 lIleGluAspMetGlnLysLysThrArgGluLeuAlaPheAlaThrGluG 1922  
852 ATCGAGGACATGCAAGAGACACGAGGAGCTGCGCTTTCGACACCGAGA 901

1922 nAspProAspAlaLysMetLeuGlnMetValLeuGlnGlySerValG 1939  
|||||  
902 GGACCCACCGATGCTAGATGCTACAGATGCTTCAGGGCTCTGTAG 951  
1939 lYProThrValAsnGlnGlyProLeuGlnValAlaGlnValPheLeuAla 1955  
|||||  
952 GGCCCAACCGTGAACCAAGGTCCCTGGAGGTGGCCAGGTGTTTATGCA 1001  
1956 GluIleProGlnAspProLysLeuPheArgHisAsnLysLeuArgLe 1972  
|||||  
1002 GAGATCCCGGAAGACCCCAAGCTCTTCGGCATCACAAATTGCGGCT 1051  
1972 uCysPheLysAspPheCys.LysLysCysGlnAspAlaLeuArgLysAsn 1988  
|||||  
1052 CTGCTTCAGGACTCTGCAAAGAAATGTGAGATGCGCTGCGGAAAAAT 1101  
1989 LysAlaLeuIleGlyProAspGlnLysGluTyrHisArgGluLeuGluAr 2005  
|||||  
1102 AAGCCCTGATGGCCGGACCAAGAGATACACCGTGAGCTGGAGCG 1151  
2005 gAsnTyrCysArgLeuArgGluAlaLeuGlnProLeuLeuThrGlnArgL 2022  
|||||  
1152 CAACCTACTGGCGCTGGGGAGGCTCTGCAGCCCTGCTTACCCAGCGCC 1201  
2022 euProGlnLeuMetAlaProThrProProGlyLeuArgAsnSerLeuAsn 2038  
|||||  
1202 TGCCCCGAGTGATGCCACCCACCCGCGCTCAGAACTCCTTGAAC 1251  
2039 ArgAlaSerPheArgLysAlaAspLeu 2047  
|||||  
1252 AGAGCAAGTTTCCGAAAGCGACACTC 1278

OM of: US-09-736-968a-2 to: Issued\_Patents\_NA:\* out\_format : pfs

Date: Oct 5, 2002 3:03 PM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

#### Command line parameters:

-MODEL=framed+.p2n.model -DEV=xlh  
-O=/cgn2\_1/USPRO.spool/US09736968/r/unat\_02102002.145334.1346/app-query.fasta\_1.2139  
-DB=Issued\_Patents\_NA -OEM=fastap -SUFFIX=rni -GAPOP=12.000  
-GAPEXT=4.000 -MIMATCH=0.100 -LOPCL=0.000 -LOPEXT=0.000  
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=bls0sum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09736968 -CGN1\_1.121 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

#### Search information block:

Query: US-09-736-968a-2  
Query length: 2047  
Database: Issued\_Patents\_NA:\*  
Database sequences: 383533  
Database length: 122816752  
Search time (sec): 113.010000

#### Score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-258-261B-6 +		221.00	297.79	6.7e-09	28958	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-456-837-6 +		221.00	297.79	6.7e-09	28958	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-457-342-6 +		221.00	297.79	6.7e-09	28958	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-457-646A-6 +		221.00	297.79	6.7e-09	28958	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-458-076A-6 +		221.00	297.79	6.7e-09	28958	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-764-233A-6 +		221.00	297.79	6.7e-09	28958	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-457-335A-6 +		221.00	297.79	6.7e-09	28958	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-729-214-6 +		221.00	297.79	6.7e-09	28958	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-028-934-6 +		221.00	297.79	6.7e-09	28958	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-764-233A-1 +		221.00	289.69	1.9e-08	49377	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-588-985-1 +		220.00	318.66	4.6e-10	6519	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-971-988-1 +		220.00	318.66	4.6e-10	6519	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-923-137-2 +		191.00	241.32	9.4e-06	36519	
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-804-227C-7 +		182.00	222.48	0.0001	44377	
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-804-198-1 +		182.00	222.48	0.0001	44377	
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-09-147-236-1 +		177.00	228.37	4.9e-05	16836	
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-09-147-236-10 +		177.00	228.37	4.9e-05	16836	
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-09-144-085-3 +		173.00	210.85	0.0005	33529	
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-09-443-501A-2 +		163.50	182.49	0.0177	71989	
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-09-103-840A-1 +		162.50	118.23	6.7e-08	4411529	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-07-689-008-1 +		160.50	207.88	0.0007	9540	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-428-517-1 +		158.50	178.92	0.0280	50937	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-320-878-19 +		154.50	176.11	0.0402	38506	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-335-409-1 +		154.50	167.30	0.1243	68750	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-568-102-1 +		154.50	167.30	0.1243	68750	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-568-969-1 +		154.50	167.30	0.1243	68750	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-568-480-1 +		154.50	167.30	0.1243	68750	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-568-486-1 +		154.50	167.30	0.1243	68750	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-568-472-1 +		154.50	167.30	0.1243	68750	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-368-550-1 +		153.00	197.68	0.0025	7812	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-171-461-1 +		153.00	174.61	0.0726	43804	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-335-409-1 +		153.00	164.66	0.1746	68750	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-568-102-1 +		153.00	164.66	0.1746	68750	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-568-969-1 +		153.00	164.66	0.1746	68750	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-568-480-1 +		153.00	164.66	0.1746	68750	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-568-472-1 +		153.00	164.66	0.1746	68750	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-320-878-21 +		152.50	212.45	0.0004	2787	
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-125-468-1 -		152.50	176.37	0.0389	30001	

/cgn2\_6/ptodata/2/ina/5B.COMB.seq:US-08-474-933-1 - 152.50 176.37 0.0389 30001  
/cgn2\_6/ptodata/2/ina/5B.COMB.seq:US-08-752-760A-1 - 152.50 173.99 0.0527 35081  
/cgn2\_6/ptodata/2/ina/5B.COMB.seq:US-09-443-501A-2 + 152.50 163.08 0.2139 71989  
/cgn2\_6/ptodata/2/ina/5B.COMB.seq:US-09-029-603-4 + 152.00 199.71 0.0019 6085  
/cgn2\_6/ptodata/2/ina/5B.COMB.seq:US-08-804-227C-1 + 152.00 169.92 0.0889 43280

seq\_name: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:US-08-258-261B-6

#### seq\_documentation\_block:

Sequence 6, Application US/08258261B  
Patent No. 5639949  
GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James M.  
APPLICANT: Beck, James Joseph  
APPLICANT: Hill, Dwight Steven  
APPLICANT: Ryals, John Andrew  
APPLICANT: Gaffney, Thomas Deane  
APPLICANT: Lam, Stephen Ting  
APPLICANT: Hammer, Phillip E.  
TITLE OF INVENTION: Genes for the synthesis of  
TITLE OF INVENTION: antipathogenic substances  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ciba-Geigy Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/258,261B  
FILING DATE: 08-JUN-1994  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/457,205  
FILING DATE: 01-JUN-1995

ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8614  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28958 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

US-08-258-261B-6  
alignment\_scores:  
Quality: 221.00 Length: 1480  
Ratio: 0.350 Gaps: 78  
Percent Similarity: 42.703 Percent Identity: 22.500

alignment\_block:  
US-09-736-968a-2 x US-08-258-261B-6 ..  
Align seg 1/1 to: US-08-258-261B-6 from: 1 to: 28958

16 ThrValAlaAlaGluValAlaArgGlnValSerArgGluArgSerGlySe 32





551 LeuTyrValITyrProHisSerLeuAspHisSerArgIn...GlySer 566  
::: ||||| ||| :::: ||::: |||  
11019 CCGCAGCGCTACCTCCGGCCGACCAACATCCCTGCTGGCCCGCCGCTGC 11068  
567 ValArgAsnLeuAlaValArgValGlnTyrMetThrGlnGlyAsnProSe 583  
||| ||||| :::: :::: ||| |||||  
11069 CCTTCGCCGACCGCGGTGGCTTCTCT...TTACAGGGCGGCTCTCCCTC 11115  
583 rGln.....AlaLeuPro..... 587  
||| ||||| ||||| |||||  
11116 GCAGAGCACCCGTGGCTCGAAGGGCCATCCGCTTCGGCACACCATCT 11165  
588 .....ValIlePheGlyLysSerSerCysSerGlnPheThrArgLys 601  
||| ||||| ||||| ||||| ||||| |||||  
11166 ACCGGGACCGGGCTTCTTCGACTCGCCCTGACAGTCCGCCACCGCGCTCG 11215  
602 AlaPheThrProValValTyrHisAsnLysSerProGlnPheTyrGln 618  
||| ||||| ||||| |||||  
11216 GCCTCGACACCGTC.....GAAGA 11234  
618 uPheLysLeuHisLeuProAlaCysValThrGlnAsnHisLeuLeu 635  
||:: ||||| ||||| ::::: ::::: |||||  
11235 GCTCAGCTCGAGGGCCCTTCGCTCTCCATCGAGGACACCGCTCTC. 11283  
635 heThrPheThrHisValSerCysGlnProArgProGlyThrAlaLeuGln 651  
||::: |||||  
11284 .....CTCAG 11289  
652 ThrProValGlyPheThrTyrPleProLeuGlnHisGlyArg.... 666  
||| ||||| ||||| |||||  
11290 ATCTCGTCGGG.....CCCGTGGACGCGACGACGCAAGGCGC 11327  
667 .....LeuArgThrGlyProPhe. 672  
||| ||||| ||||| |||||  
11328 GCTCTCTTCCATPACCAGCAAGAGAGCGGCTTAGAGATGGCCCTTGA 11377  
673 .....CysLeuProVal 676  
||| |||||  
11378 CTGCGACGCGCAGCGGCTCTCTCGCGGAGCCACATCCCTCTCCGCGC 11427  
677 SerValAspGlnProProSerProSerTyrSerValLeuThrProAspVal 693  
::::: ||||| ||||| ||||| |||||  
11428 GATCTCACAGATGCGCTCCCTCGAGTGCATC.....CGGTTGA 11468  
693 AlaLeuProGlyMetArgTyrValAspGlnHisLysGlyVal...PheSer 709  
||| ||||| ||||| |||||  
11469 CCTCGAAGGCGCTPACGCAACCTGCGCAACTCGGCTTGCTTCAGGCC 11518  
709 aGlnLeuThrAlaValSerSerValHisProGlnAspProTyrLeuAsp 725  
||| ||||| ::::: ||||| ||||| |||||  
11519 CCGAATTCAGAGGCGCTCGGCTCGCTCTACAAACGGGAGCAGAGCTTT 11568  
726 LysPhePheThrLeuValHisValLeuGlnGlnGlyAlaPheProPhe 742  
||| ||||| ||||| ||||| |||||  
11569 GCGGAGCCCAAGCTCCCGGAAAGCGGCCCAAAAGATGCCCGCGTTTGC 11618  
742 gLeuLys.....AspThrValLeuSerGlnGlyAsnValGln 755  
||| ||||| ||||| ||||| |||||  
11619 CCTCACCCCTGGCTGCTGCACAGCGCCCTGATCAGTGGCTTTGAGG 11668  
755 InGlnLeuArgLysLeuAla.....Ala 763  
::: ||||| ||||| ||||| |||||  
11669 ACGAGAGAGAGGAGCGGCTGCTGCCCTTCTCGTGAACGGAGTCTCG 11718  
764 LeuArgLeuAlaSerProGlnProLeuValAlaPheSerHisHisValle 780  
||| ||||| ::::: ||||| ||||| |||||  
11719 CTGGCGTCCGCGGCGCAACCACTTGGCGCGGCTTCAC..... 11760  
780 uAspLysLeuValArgLeuValIleArgProIleIleSerGlyGln 797  
||| ||||| ||||| |||||  
11761 .....CGTCCCAAGGTGAATCTCCGCT 11785

[illegible]

```

12425 TCCGAGGTTGGCGCAACGAGCTCGAGCGAGCGGCTCAATCCGCCCAA 12474
1088 A1aPProAspProLysValThrSerMetPheGluLeuSerGlyProPheAr 1104
12475 GCACCCACGTCGGCCCTC..... 12492
1104 gGlnGlnHisPheLeuAlaGlyLeuLeuThrGluLeuAlaLeuAlaL 1121
12493 .....CATATTCGACCAAGGACACCTTCGACGCGCTCGCCCTCGTGC 12535
1121 euGluProGluAlaGluGlyAlaPheLeuLeuHisLysLysAlaIleSer 1137
12536 ACCGCCCCGAGGCCAG...GCGCCCTCGCACACGCGCAAGTCCGCATC 12582
1138 A1aValHisSerLeu..... 1142
12583 GCGGTGACGCGGAGGCTCAACTTCGCGGATGTCGTGCACACCTTGG 12632
1143 .....LeuCysGlyHis..... 1146
12633 CATGTATCCGGCGAGCGCGCCGCTCGAGGCGCAAGCGCGGCATCG 12682
1147 ....AspThrAspProArgTyrAlaGluAlaThrValLysAlaArgVal 1161
12683 TTAAGTGAAGTCGGTCCAGGTGTCTCCGATACACGTAAGCGACGCGGTG 12732
1162 ..... 1163
12733 ATGGGGCTTTCGGCGGAGCCTTGGTCCACGCGCATCGCCGCGCCGCG 12782
1163 uLeuTyrLeuProLeuLeuSerIleAlaArgAspThrLeuProArgLeuH 1180
12783 CATGATCTGCCCC.....ATCCCCACGCGCT 12808
1180 IsaAspPheAlaGluGly.ProGlyGlnArgSerArgLeuAla...SerMe 1195
12809 GGTCTCTGCGCCAAAGCCGCGGCTCCCAATCATATCTCACCGCCTAC 12858
1195 tLeuAspSerAspThrGluGlyLysAlaGlyThrIleAspN 1212
12859 TATGGAGCTCGTCATCTCGGCG.....ATCTGAACCCCATATCAGC 12899
1212 roserValAlaMet.....AlaIleAlaGlyLysProLeuAlaPro 1225
12900 TGTCTCTCATTCATCGCGCCGCGCGCGCTCGGAGCGCGCTTCAGC 12949
1226 GlySerArgAlaSerIleSerGlnGlyProProThrAlaSerArgAlaG 1242
12950 TCCGACGCCACCTCGCGCGCGAGGCTTTCGCCACCGCCAGTCCAGGGAAG 12999
1242 yCysAlaLeuSerAlaGluSerSerArgThrLeuLeuAlaCys..... 1256
13000 TGGAGCGCTCTCGCGCGGCTCGGCTTCGAGCATGCGCAGCTCGCTC 13049
1257 ..ValLeuTrpValLeuLysAsnThrGlu.....ProAlaLeuLeuGln 1270
13050 ACCTGACCTGGGCTTCGAGCAGCATTCCTCGGCTCCACGCATGGCGCG 13099
1271 ArgTrpAlaThrAspLeuThrLeu 1278
13100 GCATGATGTCGTCCTCGACTGTC 13123
seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-456-837-6
seq_documentation_block:
; Sequence 6, Application US/08456837
; Patent No. 5643774
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven

```

```

; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,837
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-08-456-837-6

alignment_scores:
Quality: 221.00 Length: 1480
Ratio: 0.350 Gaps: 78
Percent Similarity: 42.703 Percent Identity: 22.500

alignment_block:
US-09-736-968a-2 x US-08-456-837-6 ..
Align seg 1/1 to: US-08-456-837-6 from: 1 to: 28958

16 ThrValAlaIaGluValArgLysGlnValSerArgLysArgSerGly 32
9487 ACATCGGCATCGACGTC.....GTTGATACCCCGAGAGTTCGGCG 9527
32 rProHis.....SerSerArgArgCysSer...SerSerLeuGly 45
9528 CCGCCACGACAGATAGACATCCCGCGCGCGGCGGTCTCGCCCTTCGAC 9577
45 a1ProLeuThrGluValValGluProLeuAspPheGluAspValLeuLeu 61
9578 TCTCGGACCAACAGCC.....CACGTATCTCTC 9606
62 SerArgProAspAlaGluProGlyProLeuArgAspLeuValGluPh 78
9607 GAGGAGGCTCCGCGCGCTGTGCGGC..... 9633

```

```

78 eProAlaAspAspLeuGluLeuLeuLeuGluProArgGluCysArgThr 95
   .....
9634 .....GAGCCGCCACCTCACAGACG 9655
95 hrcGluProGlyIleProLys.....AspGluLysLeu 105
   .....
9656 GCTCCGAGCCGCTCCCGCGCGCTGCCCTGCTGTCGGCCAGAGC 9705
106 AspAlaGluValArgAlaValAlaGluMetTyrIleGluAspTyrVal11 122
   .....
9706 GAGCGCGCGCTCCGCGCCAGGAGGAGCG...CTCCGGACACCTCCT 9752
122 eValHisArgArgTyrGluTyrLeuSerAlaIleTyrSerProValThr 139
   .....
9753 CGCCACGAGACCTCGCCCTTATCGATGCGCTATCGACAGGCCACCA 9802
139 hrAspThrGlu.....ArgGluArgGlu 146
   .....
9803 CCGCGCGCCCTTCGAGCACCGCGCGCTCTGCGCCGCGACCGGAC 9852
147 LysGlyLeu..ProArgGluVal..... 153
   .....
9853 GAGCTCCTCTCCGCGCTCGACCTGCGCCACAGACAGCCGCCCGAG 9902
154 ....PheGluGluAspAlaSerGlyAspGluArgSerGlyProGluAsp 168
   .....
9903 CACCGTCTCGCGCGAGCGAGCCAGCAGGAGTCTGCTGCTTTC 9952
169 SerAsnAspSerArgArgGlySerGly.....SerProGluAspThr 183
   .....
9953 CTGGGCAAGGCTCGACGTGGAGAGATGGCCCTCTCCCTGCTGCTCC 10002
183 oArgSerSerGlyAlaSerSerIlePheAspLeuArgAsnLeuAlaAla 200
   .....
1003 TCGGCGGCTCTCCGCGCTCAGC.....TCGAGCATCGCAGGCGC 10043
200 sPserLeuLeu.....ProSerLeuLeuGluArg 209
   .....
1004 GCTCGCTCTCCAGCTGAGTGAAGCTGCTCGCTCGCGCGCGAGC 10093
210 AlaAlaProGluAspValAspArgArgAsnGluThrLeuArgArgGlnH 226
   .....
10094 AGGCGCGCCCTCC.....TCGACCGCGCTCGACGTCGTA 10128
226 sArgProProAlaLeuLeuThrLeuTyrProAlaProAspGluAspGlu 242
   .....
10129 CAGCGCGCCCTCTTGGCGCTCATGTCTCCCTGCGCCCTCGGGGCTC 10178
243 Ala.....ValGluArgCysSerArgProGluProProArgGluH 256
   .....
10179 GCTCGGCGTGCAGCCCGCGCTCGCGCCACAGCCAGGCGAATCG 10228
256 sPheGlyGluArgIle..LeuValLysCysLeuSerLeuLysPheGluIle 272
   .....
10229 C.....CGCGGCTCTGTCGACAGCGCTCTCTCCTCGAGAGCGCGG 10272
273 GluIleGluProIle.....Phe 278
   .....
10273 CGCATCGCGCCCTCGCGCAAGAGCGCTCACACCGTCGCGCGCAAG 10322
278 eGlyIleLeuAlaLeuTyrAspValArgGluLysLysIleSerGluA 295
   .....
10323 CGGCATGCGCGCGCTC..... 10338
295 snPheTyrPheAspLeuAsnSerAspSerMetLysGlyLeuLeuArgAla 311
   .....
10339 .....GAGCTCGCGCGCTCCGACCTCCAGACCTACCTACCTCCTCC 10377
312 HisGlyThrHisProAlaIleSerThrLeuAla.....ArgSer 324
   .....
10378 TGGGGCGACAGG.....CTCTCCACCGCGCGCTCACAGCCCGAGGC 10421

```

```

324 rAlaIlePheSerValThrTyrProSerProAspIlePheLeuValIleL 341
   .....
10422 TACCTCGTATCCGCGAGCCCGCGCTGACGCGTGCCTC..... 10464
341 ySLeuGluLysValLeuGlnGlnLysAspIleSerGluCysGlyPro 357
   .....
10465 .....GACGTCTCACCGCCACCAAGGTG 10488
358 TyrMetValLeuLysGluValAspThrAlaLysAsnLysGluLysLeuG 374
   .....
10489 TTCGCGCCAGATCGCGCTCGACCTCCACCTCCGCGCCAGATGGA 10538
374 uLysLeuArg.....LeuAlaIleGluGlnPheCysThrArgLeuGlyA 389
   .....
10539 GCGCTTCACAGACGAGCTCGCGCA..... 10563
389 rGTYArgMetProPheAlaThrPThrAlaValHisLeuAlaAsnIleVal 405
   .....
10564 .....GCTCTAGCCACATCGCT 10581
406 SerSerAlaGlyGluLeuAspArgAspSerAspSerGluGlyArgArg 422
   .....
10582 CCTCGAGCTGCGAGCTCCCTTTATTCGACCGTCACCGGCGACAGCT 10631
422 gProAlaThrPThrAspArgArgArgGlyProGluAspArgAlaSerS 439
   .....
10632 CGAC.....GCTTCG 10642
439 eArgLysAspAlaCysSerPheSerGlyPheArgProAlaThrLeuThr 455
   .....
10643 AGCTGACGCGCGCTACGTGATCGAAACCTCCGGAACCGTCTCTG... 10689
456 ValThrAsnPhePheLysGlnGluAlaGluArgLeuSerAspGluAspLe 472
   .....
10690 .....TTCTCGAGCGCGACGCGCGCTCTCGAC...GATGG 10724
472 uPheLysPheLeuAlaAspMetArg..ArgProSerSerLeuLeuArg. 487
   .....
10725 GCATCGCTCTCCGAGGTACAGCCCATCCCTGCTCAGCTCGCGCC 10774
488 ArgLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAlaPr 504
   .....
10775 TCCGCGAGACCTCGCAGCGCTCACCGCTCGATCCGCTGCTGGCTCC 10824
504 oGluAsnProHisPheCysLeuSerPro.....Glu..LeuLeuHisIle 518
   .....
10825 ATTGCAGCAGAAAGAGCCACCTCGCCGCTCTCTCTCTGCGCGGA 10874
519 LysProTyrProAspProArgGlyArg..... 527
   .....
10875 GCTCTCTACCGCAGGCTCGCGCTGACTGGAGAGACTTCTTCGCGCCT 10924
528 .....ProThrLysGluIleLeuGluPheProAlaA 538
   .....
10925 AGCTCCCCCGCAGGCTCTCCCTCCAC.....CTACCCCTTCACAGGA 10968
538 rGluValTyrAlaProHisThrSerTyrArgAsnLeu..... 550
   .....
10969 GAGCGGTTCTGCTCGAGCTCTCCACGAGCAGACGCTCCGACGTCCCT 11018
551 LeuTyrValTyrProHisSerLeuAsnPheSerSerArgGln..GlySer 566
   .....
11019 CCGAGGCTCAGCTCGCGCAGCAATCCGCTCGCGCGCGCGCTCG 11068
567 ValArgAsnLeuAlaValArgValGlnTyrMetThrGlyGluAspProse 583
   .....
11069 CTTGCGCGACCGCGCTGCTTCTCT...TTACAGGCGGCTCTCTCCTC 11115
583 rGln.....AlaLeuPro..... 587
   .....
11116 GCGAGCACCCTGCTGCTGAGAGCCATCCGCTTCGCGACACCATCTCT 11165
588 .....ValIlePheGlyLysSerSerCysSerGluPheThrArgGlu 601

```



```

1138 AlaValHisSerLeu.....LeucylslyHis..... 1142
1143 ..... 1146
12583 GCGGTGACGAGGCGACAGGCTCAACTTCCGCGATGTCGACACCCCTTGG 12632
1143 ..... 1146
12633 CATGATATCCGGGAGACGCGCCCGCTCGAGAGCGAAGGCGGCGCATCG 12662
1147 .... AspThrAspProArgTyrAlaGluAlaThrValIysAlaArgVal 1161
12683 TTTCTGAATCGGTTCAGAGTGTCTCCCGATACACCGTATAGCCACCGGGTGG 12732
1162 ..... 1163
12733 ATGGGGGCTTCGGCGACGACCTTTGGTCCACGCGCATCGCGGCGCGCG 12782
1163 uLeuTyrLeuProLeuLeuSerIleAlaArgAspThrLeuProAlaGlu 1180
12783 CATGATCTGCCCC.....ATCCCCACGGCT 12808
1180 IsasppheAlaGluIly.ProGlyAlaArgSerArgLeuAla...SerMe 1195
12809 GGTCTTCCGCGCAAGCGCGGAGGCTCCCATCATCTCATACCGGCTTAC 12858
1195 IleuAspSerAspThrGluGlyGluIlyAspIleAlaGlyThrIleAsp 1212
12859 TATGACTGCTGATCTCGAGC.....ATCTGAACCCATCATCAG 12899
1212 roSerValAlaMet.....AlaIleAlaGlyIlyProLeuAlaPro 1225
12900 TGTCTCATCATCATGCGCGCGCGCGCGCTCGGAGCGCGCGCTTACAC 12949
1226 GlySerArgAlaSerIleSerGlnGlyProProThrAlaSerArgAlaG 1242
12950 TCCGACGCCACCTCGCGCGCGAGCTTGTCCACGCCGACGTCACGAGG 12999
1242 yCysAlaLeuSerAlaGluSerSerArgThrIleuAlaCys..... 1256
13000 TGGAGGCGCTTCGCGCGCTCGAGAGATCGCGACCTCGCGCTC 13049
1257 ..ValleuThrValLeuIlyAsnThrGlu.....ProAlaLeuLeuGln 1270
13050 ACGTGACCTGGGCTTCGAGCAGACATTCCTCGCGCTCACGATGGCGGG 13099
1271 ArgTrpAlaThrAspLeuThrIleu 1278
13100 GCATGATGTCGTCTCGACTGTC 13123
seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-457-342-6
seq_documentation_block:
; Sequence 6, Application US/08457342
; Patent No. 5662898
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Unkes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA

```

```

; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,342
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 424
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; US-08-457-342-6

alignment_scores:
Quality: 221.00 Length: 1480
Ratio: 0.350 Gaps: 78
Percent Similarity: 42.703 Percent Identity: 22.500

alignment_block:
US-09-736-968a-2 x US-08-457-342-6 ..
Align seg 1/1 to: US-08-457-342-6 from: 1 to: 28958

16 ThrValAlaAlaGluValArgLysGlnValSerArgIleuArgSerGlyse 32
9487 ACACCTGCCATCGACGTC.....GTGATACCCCGAGGTCTTGGGC 9527
32 ProHis.....SerSerArgArgCysSer...SerSerLeuGly 45
9528 CCGCCAGCAGATGACGATCCCGCGCGCGCGCTCTCCGCTTCGGAGC 9577
45 aProLeuThrGluValValGluProLeuAspPheGluAspValLeuLeu 61
9578 TCTCCGCGACCAAGCC.....CACGTATCCTCTC 9606
62 SerArgProProAspAlaGluProGlyProLeuArgAspLeuValGluP 78
9607 GAGGAGGCTCCCGCGCGCTCTGGGC..... 9633
78 eProAlaAspAspLeuGluLeuLeuGlnProArgGluCysArgThrT 95
9634 .....GAGCCCGCCACCTCATACAGAGG 9655
95 hngLuproGlyLeuProLys.....AspGluLysLeu 105
9656 CCGCCAGCAGCGCTCCCGCGGCTGCGCGTCTCTCGCGCAGGAGC 9705
106 AspAlaGluValArgAlaAlaValGluMetTyrIleGluAspTrpValI 122
9706 GAGGCGCGCTCCCGCGCGCGAGCGAAGCG...CTCCGCGACCACTCTCT 9752
122 eValHisArgArgTyrGlnTyrLeuSerAlaAlaTyrSerProValThrT 139

```

```

9753 CGCCGACGAGACCTGCGCCCTTATGATGGCTATTGCGAGGCCACCA 9802
139 hrAspThrGln.....ArgGluArgGln 146
9803 CCGCGGCCCACTTCGAGACCGCGCGCTCTCTGGCCCGGACCGCGAC 9852
147 LysGlyLeu, ProArgGlnVal..... 153
9853 GAGCTCTCTCCGCGCTCGCATGCTCGCCACGACACCGCGCCGAG 9902
154 ....PheGluGlnAspAlaSerGlyAspGluArgSerGlyProGluAsp 168
9903 CACCGTTCTCGCGCGAGCGAGCCAGCGCATGCTCTGCTTTC 9952
169 SerAsnAspSerArgArgGlySerGly.....SerProGluAspThrPr 183
9953 CTGGGCAAGGCTCGCAGTGGGAAGGATGGCCCTCTCCCTCGCTCAGCTCC 10002
183 oArgSerSerGlyAlaSerSerThrPheAspLeuArgAsnLeuAlaAla 200
10003 TTCGCGGCTCTCCGCGCTCAGC.....TCGAAGCATGCGAGCGCGC 10043
200 spSerLeuLeu.....ProSerLeuLeuGluArg 209
10044 GCTCGCTCTCAGCTGAGTGAAGCTGCTGCGCTGCGCGCGAGC 10093
210 AlaAlaProGluAspValAspArgArgAsnGluThrLeuArgArgGlnH 226
10094 AGGCGCGCCCGCTCC.....TCGACCGCGCTCAGCATGCGA 10128
226 sArgProProAlaLeuLeuThrLeuThrProAlaProAspGluAspGln 242
10129 CAGCCCGCGCTCTTGGCTGCTATGGTCTCTCCGCGCGCTGCGCGCTC 10178
243 Ala.....ValGluArgCysSerArgProGluProArgGlnH 256
10179 GCTCGCGCTCGAGCCCGCGCGCTGCTGCGCCACAGCGAGCGAGATCG 10228
256 sPheGlyGlnArgGlyLeuValLysCysLeuSerLeuLysPheGluLe 272
10229 C.....CGCGCGCTTCGTGCGAGGCGCTCTCCCTCGAGGACCGCGG 10272
273 GluIleGluProIle.....Ph 278
10273 CGCATGCGCGCTGCGCGAGAAAGCGTCCACACCGTCCGCGGACAGG 10322
278 eGlyIleLeuAlaLeuThrAspValArgGluLysLysIleSerGluA 295
10323 CCGCATGCGCGCGCTC..... 10338
295 snPheThrPheAspLeuAsnSerAspSerMetLysGlyLeuLeuArgAla 311
10339 .....GAGCTCGCGCGCTCGAGCTCCAGACCTACCTCGCTCC 10377
312 HisGlyThrHisProAlaIleSerThrLeuAla.....Argse 324
10378 TGGGGGAGACAG.....CTTCACCGCGCGCTCGACAGCGCCCAAGGCG 10421
324 rAlaIlePheSerValThrThrProSerProAspThrPheLeuValIleL 341
10422 TACCTCGATCGCGGAGCGCGCGCGCTGAGCGCGCTGCTC..... 10464
341 yLeuGluLysValLeuGlnGlnGlyAspIleSerGluCysGluPro 357
10465 .....GAGCTCTCACCAGCCCAAGAGTGG 10488
358 TyrMetValLeuLysGluValAspThrAlaLysAsnLysGluLysLeuG 374
10489 TTCGCGCGCAAGATCCGCGTGCATACGCTCCCTCCGCGCAAGATGGA 10538
374 uLysLeuArg.....LeuAlaAlaGluGlnPheCysThrArgLeuGlu 389
: .....1111111111

```

---

```

10539 CGCCGTCACAGACGAGTCCGCCGA..... 10563
389 rGlyArgMetProPheAlaThrPheAlaValHisLeuAlaAsnIleVal 405
10564 .....GGTCTAGCCAAATCGCT 10581
406 SerSerAlaGlyGlnLeuAspArgAspSerAspSerGluGluArg 422
10582 CCGCGAGCTGCGAGAGTCCCTTTATTCGACGCTCACCGGACACAGGCT 10631
422 gProAlaThrPheAspArgArgArgGlyProGluAspArgAlaSer 439
10632 CGAC.....GGCTCCG 10642
439 eArgLysAspAlaCysSerPheSerGlyPheArgProAlaThrLeuThr 455
10643 AGCTCGAGCGCGCTACTGATGAAACCTCCGGCAACCGTCTC... 10689
456 ValThrAsnPhePheLysGlnGluAlaGluArgLeuSerAspGluAspLe 472
10690 .....TTCGAGCGCGACCGAGCGCTCTCGAC...GATGG 10724
472 uPheLysPheLeuAlaAspMetArg...ArgProSerSerLeuLeuArg 487
10725 GCATCGCTCTCGTGCAGAGTCAAGCCCAATCCGCTCAGCTCGGCC 10774
488 ArgLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAlaPr 504
10775 TCAGCGAGACCTGCGAGCGCTCACCGCTCGATCCCGTCTGTCGCTCC 10824
504 oGluAsnProHisPheCysLeuSerPro.....GluLeuLeuHisIle 518
10825 ATTGAGAGAGAAGAACCACTCGCGCGCTGCTCTCTCTGAGGCGGA 10874
519 LysProGlyProAspProArgGlyArg..... 527
10875 GCTCTTACCGAGGCTCGCGCTGAGTGAAGAGACTTTCGCGGCT 10924
528 .....ProThrLysGluIleLeuGluPheProAla 538
10925 ACGTCCCGCGAGAGTCCCTCCCAAC.....CTACCCCTTCAGAGCA 10968
538 rGluValThrAlaProHisThrSerThrArgAsnLeu..... 550
10969 GAGCGGTTCTGGCTGAGCTGTCACAGAGCAAGAGTTCGAGAGCGCT 11018
551 LeuThrValThrProHisSerLeuAsnPheSerSerArgGln...GlySer 566
11019 CCGCAGGCTGAGCTCGCGGACCAATCCGCTGCTGCGCGCGCTCG 11068
567 ValArgAsnLeuAlaValArgValGlnTyrMetThrGlyGluAspProse 583
11069 CTTTCGCGCGAGCGCGGTGGCTTCTCT...TTACAGGCGCGCTCTCC 11115
583 rGln.....AlaLeuPro..... 587
11116 GCAGAGCAGCCGTGGCTGAGAGGCAATGCTCTGCGCACACCATCT 11165
588 .....ValIlePheGlyLysSerSerGlySerGluPheThrArgGln 601
11166 ACCGGGACAGCGGCTTCTGAGCTCGCGCTGCGCACAGTCCGCGCGCT 11215
602 AlaPheThrProValValThrHisAsnLysSerProGluPheThrGluG 618
11216 GCTTCGACACCGCTC.....GAAAG 11234
618 uPheLysLeuHisLeuProAlaCysValThrGluAsnHisIleLeuLeu 635
11235 GCTCACGCTGAGAGCGGCTGCTGCTCCATCGAGAGACAGCGCTCTC 11283
635 hethrPheThrHisValSerCysGlnProArgProGlyThrAlaLeuGlu 651
11284 .....CTCGAG 11289

```

652 ThrProValGlyPheThrTrpIleProLeuLeuGlnHisGlyArg..... 666  
|||||  
11290 ATCTCCGTGGG.....CCGTGGACGACGACGACGACGAGGC 11327  
667 .....LeuArgThrGlyProPhe. 672  
11338 GCTCTTTTCATAGACGACGACGACGACGCTTCAGATGATGCCCCCTGGA 11377  
673 .....CysLeuProVal 676  
11378 CTCGGCAGCGCAGCGCTCTCTCCGCGCGGACCCATCCCTCTCCGCC 11427  
677 SerValAspGlnProProSerTyrSerValLeuThrProAspValAl 693  
: : : : :  
11428 GATCTCCACGAGTGGCTCCCTCGAGTGCATC.....CCGGTGA 11468  
693 aLeuProGlyMetArgTrpValAspGlyHisGlyVal...PheSerV 709  
|||  
11469 CTTGAAAGGCTCTACGACACCTCGCCACCTGGGCTTGCCTACGAGCC 11518  
709 aGluLeuThrAlaValSerSerValHisProGlnAspProTyrLeuAsp 725  
|||  
11519 CCGAGTTCAGAGGCTCCGCTCCGTACAGAGCGCGGACGACGACTTTT 11568  
726 LysPhePheThrLeuValHisValLeuGluGluGlyAlaPheProPheAr 742  
|||  
11569 GCCGAGCGCAAGCTCCGGAAGCGCGGAAAGGATGCCGCCGCTTTCG 11618  
742 gLeuLys.....AspThrValLeuSerGluGlyAsnValGlu 755  
|||  
11619 CCTCACCCCTCGCTGCTCGACGACGCGCTTCATGCACTGGCTTTAGG 11668  
755 InGluLeuArgAlaSerLeuAla.....Ala 763  
|||  
11669 ACGAGCAGAGGAGCGATCCCTCTGCTTCGTGAGGAGGAGTCTCG 11718  
764 LeuArgLeuAlaSerProGluProLeuValAlaPheSerHisValLe 780  
|||  
11719 CTGGGCTCCGTGGTCCACCACTTCGGGTGGCTTCAC..... 11760  
780 uAspLysLeuValArgLeuValIleArgProProIleIleSerGlyInI 797  
|||  
11761 .....CGTCCCAAGGATGATCCCTCTCT 11785  
797 LeuValLeuGlyValGlyAlaPheGluAlaMetAlaHisValValSer 813  
: : : : :  
11786 CGATCGTCTGGCGGACGCGCAGTGTGCTTCGCTGCGTGCAGGCG 11835  
814 LeuValHisArgSerLeuGluAlaAlaGln.....AspAlaArgG 827  
|||  
11836 CTCGCCATGGGACGACGCTCGCGCGGACGCTCGCACCCCGGAGCTTC 11885  
827 yHisCysProGlnLeuAlaAlaTyrValHisTyrAlaPheArgLeuPro 844  
|||  
11886 CCACCATGATGCGCTC.....TTCGGGTGCAC 11914  
844 LyrThrGluProSerLeuProAspGlyAlaProProValThrValGlnAla 860  
: : : : :  
11915 GAGCGAGCTCCAAAGGCCCACT...TCACCGCTCGCGCCCGGAGGCGC 11961  
861 AlaThrLeuAlaArgGlySer.....G 868  
: : : : :  
11962 GTCTTTTCGGCACAGCGGCGCAGATCTCGGCTGACGCCCGCTCGC 12011  
868 yArgProAlaSerLeuTyrLeuAlaArgSerLys...SerIleSerSer 884  
: : : : :  
12012 CCGCTACGCGACCTCGCTCGCGAAGCGCGCTCGACCAAGGCGCTT 12061  
884 eArgnProAspLeuAlaValAlaProGlySerValAspAspGlyValSer 900  
|||  
12062 GCGCTCCGCGCTCGCTGCTGCGCCCT...TTCATGATCGACCGGACGAGC 12108

901 ArgIleuAlaSerLysLeuLeuHisGluGluLeuAlaLeuGlnTrpVal 917  
: : : : :  
12109 GACCTGCTCCCGAC.....GCCACAGAC..... 12132  
917 ValSerSerSerAlaValArgGluAlaIleLeuGlnHisAlaTrpPhe 934  
: : : : :  
12133 .....GCCACGCGCTCGCACTCGCCCTTCGAA...GCTGG.... 12168  
934 hePheGlnLeuMetValLysSerMetAlaLeuHisLeuLeuGlyGln 950  
|||  
12169 .....CTCGCGGACGAA 12180  
951 ArgLeuAspThrProArgLysLeuArgPheProGlyArgPheLeuAsp 967  
|||  
12181 GCGCTCGCCCTCGCGCGC..... 12198  
967 pIleThrAlaLeuValGlySerValGlyLeuGluValIleThrArgValH 984  
|||  
12199 .....CTGCTCTGTCACCCGACGCGCGCTCGCCACCCACACCG 12238  
984 IsLysAspValGluLeuAlaGlnHisLeuAsnAlaSerLeuAlaPhePhe 1000  
: : : : :  
12239 AAGACACGTCAG.....CTCGCCGCTTTC 12252  
1001 LeuSerAspLeuLeuSerLeuValAspArgGlyPheValPheSerLeuVal 1017  
: : : : :  
12253 .....GACCTGCC 12260  
1017 LArgAlaHisTyrGluGlnValAlaThrArgLeuGlnSer...SerProAs 1033  
: : : : :  
12261 TCAGCGCGCGCTCTGGGGCTCGCGGCTCGCGCAAACTGAGCACCCAG 12310  
1033 nProAlaAlaLeuLeuThrLeuArgMetGluPheThrArgIleLeuCys 1050  
|||  
12311 ACCTCCCGCTCT.....C 12334  
1050 eHisGlnHisTyrValThrLeuAsnLeuProCys..... 1061  
: : : : :  
12325 CTGCTGACATGCACTTCAGCGAGGCTCCAGACGAGCCCTGTAAGGCGC 12374  
1062 .....CysProLeuSerProPro.....AlaSerProse 1071  
|||  
12375 GCTCGACACAGAAAGCGCAGCTCGCTCCGCAAGGAAACCCCTCA 12424  
1071 rProSerValSerSerThrThr...SerGlnSerThrPheSerSerGln 1087  
|||  
12425 TCCCGAGTTGGGCAACACGCTCGACGAGCGGCTATCCGCGGCA 12474  
1088 AlaProAspProLysValThrSerMetPheGluLeuSerGlyProPheAr 1104  
|||  
12475 GCACCCACGTGGCGCTC..... 12492  
1104 gGlnGlnHisPheLeuAlaGlyLeuLeuLeuThrGluLeuAlaValAl 1121  
|||  
12493 .....CATATTCGACCAAGGACACTTCGACGCGCTCGGCTCGTCG 12535  
1121 euGluProLualagluGlyAlaPheLeuLeuHisLysLalIleSer 1137  
|||  
12536 ACGCCGCCGAGGCCAG...CGCGCTTCGCACAGGCAAGTCCGATC 12582  
1138 AlaValHisSerLeu..... 1142  
|||  
12583 GCGGTGCAGCGGCAAGGCTCACTTCGCGATGTGTCGACACCCCTTG 12632  
1143 .....LeuCysGlyHis..... 1146  
|||  
12633 CATGTATCGGCGGACGCGCGCTCGGAGGCGCAAGCGCGGCGATCG 12682  
1147 .....AspThrAspProArgTyrAlaGluAlaThrValLysAlaArgVal 1161  
: : : : :  
12683 TTAGTAAGTGGTCCAGGTGTCTCCGATACACCGTACGCGGCGGTG 12732  
1162 .....AlaG 1163

```
12733 ATGGGGCTCTGGCGCCAGCCTTTGGTCCAGCGCCATCGCCGAGCCCG 12782
1163 uleuYrleuPProLeuLeuSerlleAlaIArgspThreuProArgleuH 1180
12783 CATGATCTGCCCC.....ATCCCCCAGCGCT 12808
1180 lAspPheAlaIguLy ProGlyInArgSerArgleuAla...Serle 1195
12809 GGTCTTTCGCCAGCGCCAGGCTCCATCATCTATCTACCGCCCTTC 12858
1195 tleuAspSerArgThrluGlyluGlyaspIleAlaIguThrlleuAsp 1212
12859 TATGAGACTGTCGATCTCGGCG.....ATCGAAGAACCAATCAACG 12899
1212 roSerValAlaMet.....AlaIleAlaIguLyProleuAlaPro 1225
12900 TGTCTTCATCCATCGCGCCGCGCGCGCTCGGAGCGCCGCTTCAGC 12949
1226 GlySerArgAlaSerlleSerGInLyProProThrAlaSerArgAlaI 1242
12950 TCCACAGCCACCTCGCGCCGAGGTCTTGCACCGCCAGTCCAGGAG 12999
1242 yCysAlaLeuSerAlaIguSerArgThrlleuAlaCys..... 1256
13000 TGGAGCGCTCTCCGCGCTCGCTCGACGATGCGCACCTCGCGTCTC 13049
1257 .ValleuTrpValleuLyAsnThrlu.....ProAlaLeuGIn 1270
13050 ACCTGACCTCGGCTTCGAGCAGCATCTCGCTCGACGATGGCGCG 13099
1271 ArgTrpAlaThraspleuThrlu 1278
13100 GCATGGATGCTGCTCGCTCGACTGTC 13123

seq_name: /cgn2.6/ptodata/2/ina/5A-COMB.seq:US-08-457-646A-6

seq_documentation_block:
; Sequence 6, Application US/08457646A
; Patent No. 5679560
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Unnes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,646A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
```

```
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8689
; TELEFAX: 919-541-8614
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-08-457-646A-6

alignment_scores:
Quality: 221.00 Length: 1480
Ratio: 0.350 Gaps: 78
Percent Similarity: 42.703 Percent Identity: 22.500

alignment_block:
US-09-736-968A-2 x US-08-457-646A-6 ..
Alignment 1/1 to: US-08-457-646A-6 from: 1 to: 28958

16 ThrValAlaAlaIguValArgLyGInValSerArgIuArgSerGly 32
|||||.....
9487 ACACTCGCATCGACGTC.....GTGATACCCCGAGGCTTGTGGC

32 rProHis.....SerSerArgGlySer...SerSerLeuGly 45
|||||.....
9528 CGCCACGACAGATAGCACTCCCGCGCGCGCTCGCTCCGCTTCGAC

45 alProleuThrluValValGluProleuAspPheGluAspValLeu 61
|||||.....
9578 TCTCGGACCAACGCGC.....CACGTATCTCTC

62 SerArgProProAspAlaIguProGlyProleuArgAspLeuValGlu 78
|||||.....
9607 GAGGAGCTCCCGCGCGCTGCGGCG.....

78 eProAlaAspAspLeuGluLeuLeuGInProArgGluCysArgThr 95
|||||.....
9634 .....GAGCCGCGCACCTCACAGACG 9655

95 hrGluProGlyIleProLy.....AspGluLyLeu 105
|||||.....
9656 GCTCGCAGCGCTCCCGCGCGGTGTGCGGCTCTGTCGCGCAGAGC 9705

106 AspAlaIguValArgAlaAlaValGluMetYrIleGluAspTrpVal 122
|||||.....
9706 GAGCGCGCGCTCCCGCGCGCGCGAGCG...CTCGGACACACCTCT 9752

122 eValHisArgArgYrGInYrleuSerAlaAlaYrSerProValThr 139
|||||.....
9753 CGCCACGACGACGCTCGCGCTTATCGATGTGCTATCGACGACGAC 9802

139 hrAspThrlu.....ArgGluArgGln 146
|||||.....
9803 CCGCGCGCGCTTCGAGCAGCGCGCGCTCTGCGCGCGAGCGGAC 9852

147 LysGlyLeu.ProArgGluVal..... 153
|||||.....
9853 GAGCTCTCTCGCGCTGACTCGCTCGCCAGACAGACCGCGCGAG 9902

154 .....PheGluGlnAspAlaSerGlyAspGluArgSerGlyProGlu 168
|||||.....
9903 CACGTTCTCGCGCGCGAGCGAGCGACGACGAGTGTCTTCTCTTTC 9952
```





```

11428 GATCTCCAGAGTGCCCTCCCTGAGTGCATC.....CGGGGA 11468
      :::::  :::::  |||||  :::::  ||
693  AleuProgluMetArgTrpValAspGlnHisLysGlyVal...PheSerV 709
      |||  |||||  :::::  :::::  |||||  :::::
11469 CCTCGAAGCCCTTACGCAAGCCCTCCGCAACCTCGGGCTTCCCTACGGCC 11518
      |||||  :::::  |||||  :::::  :::::
709  AlGluLeuThrAlaValSerSerValHisProGlnAspProTyrLeuAsp 725
      |||||  :::::  |||||  :::::  ||
11519 CCGAGTTCAGAGGCTCCGCTCCCTCTACAAAGCCGGCGAGAGCTCTTT 11568
      |||||  :::::  |||||  :::::  ||
726  LysPhePheThrLeuValHisValLeuGlnGluValAlaPheProPheAr 742
      |||  :::::  |||||  :::::  ||
11569 GCGCAAGCCAGCTCCCGAAGCGCGCAAAAGATGCGCGCGCTTGGC 11618
      |||  :::::  |||||  :::::  ||
742  GlnLys.....AspThrValLeuSerGluGlnValGlnG 755
      |||  |||||  :::::  |||||
11619 CCTCCACCTCGCTCGTGCAGAGCGCTCGATGCACCTGGCTTTGAGG 11668
      |||||  :::::  |||||  :::::
755  InGluLeuArgAlaSerLeuAla.....Ala 763
      |||||  |||||  :::::  |||||
11669 ACGACAGAGAGGAGCGGTGCTGCGCCCTTCGTGAGCGAGTCTCG 11718
      |||||  :::::  |||||  :::::
764  LeuArgLeuAlaSerProGluProLeuValAlaPheSerHisValLe 780
      |||||  :::::  |||||  :::::
11719 CTGGCTCTCGTGGTGCACCACTTGGCGCTTGCAC..... 11760
      |||||  :::::  |||||  :::::
780  uAspLysLeuValArgLeuValIleArgProIleIleSerGlnI 797
      |||||  |||||  :::::  |||||
11761 .....GTCGCCAAGGGTGAATCTCCGCT 11785
      |||||  :::::  |||||  :::::
797  IValAsnLeuGluArgGlyAlaPheGluAlaMetAlaHisValSer 813
      :::::  |||||  :::::  |||||
11786 CGATGCTCTGGCGAGCGCGAGGTGACCTTGGCTGGTGCAGAGCG 11835
      |||||  :::::  |||||  :::::
814  LeuValHisArgSerLeuGluAlaAlaGln.....AspAlaArgG 827
      |||||  :::::  |||||  :::::
11836 CTGGCGAAGCGAGAGCTCGCGCGAGCTCGCAACCCCGCAGCTTC 11885
      |||||  :::::  |||||  :::::
827  uHisCysProGlnLeuAlaAlaTyrValHisTyrAlaPheArgLeuPro 844
      |||||  |||||  :::::  |||||
11886 CCACCATATGCGCTC.....TTCCGCTCGACT 11914
      |||||  :::::  |||||  :::::
844  LyrThrGluProSerLeuProAspGlyAlaProProValThrValGlnAla 860
      :::::  |||||  :::::  |||||
11915 GGAGCGAGCTCCAAAGCCCACT...TCACGCTCGCGCCGAGGAGGC 11961
      |||||  :::::  |||||  :::::
861  AlaThrLeuAlaArgGlySer.....G 868
      :::::  |||||  :::::  |||||
11962 GTCTTCTGCGACAGCGCGCAGATCTCGGCTCGAGCCCGCTGCG 12011
      |||||  :::::  |||||  :::::
868  uArgProAlaSerLeuTyrLeuAlaArgSerLys...SerLiesers 884
      |||||  |||||  :::::  |||||
12012 CCGCTACGCGCAGCTCGCTGCTCCGAAAGCGCCTCGACCAAGGCGCTT 12061
      |||||  :::::  |||||  :::::
884  eAsnProAspLeuAlaValAlaProGlySerValAspAspGluValSer 900
      |||||  |||||  :::::  |||||
12062 CGCCTCCGCGCTCGTCTGCGCCCC...TTCATGATCGACCGGAGGC 12108
      |||||  :::::  |||||  :::::
901  ArgIleLeuAlaSerLysLeuLeuHisGlnGluLeuAlaLeuGlnTrpVa 917
      |||||  |||||  :::::  |||||
12109 GACCTGTCGCGAGC.....GCCACAGC..... 12132
      |||||  :::::  |||||  :::::
917  lValSerSerSerAlaValArgGluAlaIleLeuGlnHisAlaTrpPheP 934
      :::::  |||||  :::::  |||||
12133 .....GCCACCGGCTCGCACTCGCCCTTGCAG...GCTGG.... 12168
      |||||  :::::  |||||  :::::
934  hepHeGlnLeuMetValLysSerMetAlaLeuHisLeuLeuLeuGln 950
      |||||  |||||  :::::  |||||
12169 .....CTCGCGGAGAA 12180
      |||||  :::::  |||||  :::::
951  ArgLeuAspThrProAlaGlyLysLeuArgPheProGlyArgPheLeuAspAs 967
      |||||  :::::  |||||  :::::

```

```

12181 CGCCTCGCCTCGTGGCGC..... 12198
      |||||  :::::  |||||  :::::
967  pIleThrAlaLeuValGlySerValGlyLeuGluValIleThrArgValn 984
      |||||  |||||  :::::  |||||
12199 .....CTGCTCTGCTGACCGGAGCGCGCTCGCCACCAACCG 12238
      |||||  :::::  |||||  :::::
984  lLysAspValGluLeuAlaGlnHisLeuAsnAlaSerLeuAlaPhePhe 1000
      |||||  |||||  :::::  |||||
12239 AAGACGACGTCAG..... 12252
      |||||  :::::  |||||  :::::
1001  LeuSerAspLeuLeuSerLeuValAlaPargGlyPheValPheSerLeuVa 1017
      |||||  |||||  :::::  |||||
12253 .....GACCTCGC 12260
      |||||  :::::  |||||  :::::
1017  lArgAlaHisTyrLysGlnValAlaThrArgLeuGlnSer...SerProAs 1033
      |||||  |||||  :::::  |||||
12261 TCACGGCGCGCTCTGGGGCTCGCGGCTCCGCGCAAAATGAGACCCAG 12310
      |||||  |||||  :::::  |||||
1033  nProAlaAlaLeuLeuThrLeuArgMetGluPheThrArgIleLeuCys 1050
      |||||  |||||  :::::  |||||
12311 ACCTCCGCTCT.....C 12324
      |||||  :::::  |||||  :::::
1050  eHisGlnHisTyrValThrLeuAsnLeuProCys..... 1061
      |||||  |||||  :::::  |||||
12325 CTGCTGACATCGACCTCAGCGAGGCTCCGACAGCCCTCTAGGCGC 12374
      |||||  |||||  :::::  |||||
1062 .....CysProLeuSerProPro.....AlaSerProSe 1071
      |||||  |||||  :::::  |||||
12375 GCTCGACAGAGAGAACGCCAGCTCGCTCCGCAAGGGAAACCCCTTA 12424
      |||||  |||||  :::::  |||||
1071  rProSerValSerSerThrThr.SerGlnSerSerThrPheSerSerGln 1087
      |||||  |||||  :::::  |||||
12425 TCCGAGGTGGGCGCAACACGCTCGACGAGCGCTCATCCCGCGCA 12474
      |||||  |||||  :::::  |||||
1088  AlaProAspProLysValThrSerMetPheGluLeuSerGlyProPheAr 1104
      |||||  |||||  :::::  |||||
12475 GCACCCACGTGGCGCTC..... 12492
      |||||  |||||  :::::  |||||
1104  gGlnGlnHisPheLeuAlaGlyLeuLeuLeuThrGluLeuAlaAlaAl 1121
      |||||  |||||  :::::  |||||
12493 .....CATATTCGACCAAGGCACTTCGAGCGGCTCGGCTCTGCTG 12535
      |||||  |||||  :::::  |||||
1121  euGluProGluAlaGluGlyAlaPheLeuLeuHisLysLysAlaIleSer 1137
      |||||  |||||  :::::  |||||
12536 ACGCCCCGAGGGCCAG...GCGCCCTCGCACAGCGCAAGTCCGCATC 12582
      |||||  |||||  :::::  |||||
1138  AlaValHisSerLeu..... 1142
      |||||  |||||  :::::  |||||
12583 GCGGTGCACGCGGACAGGCTCAACTCCGCGATGTGTCGACACCCTTG 12632
      |||||  |||||  :::::  |||||
1143 .....LeuCysGlnHis..... 1146
      |||||  |||||  :::::  |||||
12633 CATGTATCCGGGGAGCGCGCGGCTCGAGGCGAAGCGCGGAGATCG 12682
      |||||  |||||  :::::  |||||
1147 .....AspThrAspProArgTyrAlaGluAlaThrValLysAlaArgVal 1161
      |||||  |||||  :::::  |||||
12683 TTACTGAAGCGGTCCAGGTGTCTCCGATACCGCTAGGCAAGCGGGTG 12732
      |||||  |||||  :::::  |||||
1162 .....AlaG 1163
      |||||  |||||  :::::  |||||
12733 ATGGGGGTCTTGGCGCGACGCTTGGTCCCAAGCGCATGCGGAGCGCG 12782
      |||||  |||||  :::::  |||||
1163  uLeuTyrLeuProLeuLeuSerIleAlaArgAspThrLeuProArgLeuH 1180
      |||||  |||||  :::::  |||||
12783 CATGATCTGCGCC.....ATCCCGCAAGCCT 12808
      |||||  |||||  :::::  |||||
1180  iAspPheAlaGluGly.ProGlyGlnArgSerArgLeuAla...SerMe 1195
      |||||  |||||  :::::  |||||
12809 GTGCTTTCGCCCAAGCGCGCAGGTCGCCCATCATATCTCAACGCGCTAC 12858
      |||||  |||||  :::::  |||||
1195  tLeuAspSerAspThrGluGlyGlnLysPheIleAlaGlyThrIleAsn 1212
      |||||  |||||  :::::  |||||
12859 TATGACTGTGTGATCTGGGC.....ATCTGAACCAACCAATCAACG 12899
      |||||  |||||  :::::  |||||

```

```

1212 roSerVala1aNet.....Ala1leAlaGlyGlyProLeuAlaPro 1225
      ||| :||| ||| :||| |||
12900 TGTCTCATCATCGCGCCGCCGCCGCCGCCGCCGCCGCCGTTCCAGC 12949
      :||| :||| :||| :||| :|||
1226 GlySerArgAlaSer1leSerGlnGlyProProThrAlaSerArgAlaG1 1242
      :||| :||| :||| :||| :|||
12950 TCGCACGCCCATCTCGCGCCGCCGCCGCCGCCGCCGCCGCCGTTCCAGC 12999
      :||| :||| :||| :||| :|||
1242 yCysAlaLeuSerAlaGluSerSerArgThrLeuAlaGly..... 1256
      :||| :||| :||| :||| :|||
13000 TCGAGCGCTCTCGCGCGCTCGCGCTCGCGCATCGCACCTCGCGCTC 13049
      :||| :||| :||| :||| :|||
1257 ..ValLeuTrpValLeuLysAsnThrGlu.....ProAlaLeuLeuGln 1270
      ||| ||| :||| :||| :|||
13050 ACGTGACCTGGGCTTCGACGACGACTCTCGCGCTCGCGCATCGAGCGCG 13099
      ||| :||| :||| :||| :|||
1271 ArgTrpAlaThrAspLeuThrLeu 1278
      ||| :||| :||| :||| :|||
13100 GCATGATGTCTCTTCGACTGTC 13123

```

seq\_name: /cgn2\_6/ptodata-a/2/lna/5A\_COMB.seq:us-08-458-076A-6

seq\_documentation\_block:

Sequence 6, Application US/08458076A

Patent No. 5698425

GENERAL INFORMATION:

APPLICANT: Schupp, Thomas

APPLICANT: Ligon, James M.

APPLICANT: Beck, James Joseph

APPLICANT: Hill, Dwight Steven

APPLICANT: Ryals, John Andrew

APPLICANT: Gaffney, Thomas Deane

APPLICANT: Lam, Stephen Ting

APPLICANT: Hammer, Phillip E.

APPLICANT: Uknes, Scott Joseph

TITLE OF INVENTION: Genes for the synthesis of

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

ADDRESS: Ciba-Geigy Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: NY

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/458,076A

FILING DATE: 01-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/457,205

FILING DATE: 01-JUN-1995

APPLICATION NUMBER: 08/258,261

FILING DATE: 08-JUN-1994

ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129

REFERENCE/DOCKET NUMBER: CGC 1506/CIP3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8614

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 28958 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

```

; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-458-076A-6

```

```

alignment_scores:
  Quality: 221.00      Length: 1480
  Ratio: 0.350         Gaps: 78
  Percent Similarity: 42.703  Percent Identity: 22.500

```

alignment\_block:

US-09-736-968A-2 x US-08-458-076A-6

Align seg 1/1 to: US-08-458-076A-6 from: 1 to: 28958

```

16 ThrValAlaAlaGluValArgLysGlnValSerArgLysArgSerGlyse 32
   |||:|||| :||| ||| ||| ||| |||
9487 ACGCTCGCCATCGACGTC.....GTGATACCCCGAGGCTTGCGC 9527
   :||| :||| :||| :||| :|||
45 alProLeuThrGluValValGluProLeuAspPheGluAspValLeuLeu 61
   :||| :||| :||| :||| :|||
9528 CCGCCACGAGATAGACAGTCGCCGCCGCCGCCGCCGCCCTTCGGAC 9577
   :||| :||| :||| :||| :|||
9578 TCTCCGGACCCACGCC.....CACGTATCTTC 9606
   :||| :||| :||| :||| :|||
62 SerArgProProAspAlaGluProGlyProLeuArgAspLeuValGluPh 78
   :||| :||| :||| :||| :|||
9607 GAGAGAGCTCCCGCGCGCGCTGCGGC..... 9633
   :||| :||| :||| :||| :|||
78 eProAlaAspAspLeuGluLeuLeuGlnProArgGluLucArgThrT 95
   :||| :||| :||| :||| :|||
9634 .....GAGCCGCCACCTCACAGACG 9655
   :||| :||| :||| :||| :|||
95 hrgLupProGlyIleProLys.....AspGluLysLeu 105
   :||| :||| :||| :||| :|||
9656 CGTCGACCGCTCCCGCGCGGTGTCCTGTCGCGCCGACGAGC 9705
   :||| :||| :||| :||| :|||
106 AspAlaGlnValArgAlaAlaValGluMetLysTrpValAla 122
   :||| :||| :||| :||| :|||
9706 GAGGCGCGCTCCCGCGCGCGCGAGCAGCGG...CTCGGCACACCTCCT 9752
   :||| :||| :||| :||| :|||
122 eValHisArgArgTrpGlnTrpLeuSerAlaAlaTrpSerProValThr 139
   :||| :||| :||| :||| :|||
9753 CGCCACGACGACCTCGCCCTTATGATGCTATTCGAGCGCACCA 9802
   :||| :||| :||| :||| :|||
139 hraspThrGln.....ArgGluArgGln 146
   :||| :||| :||| :||| :|||
9803 CCGCGGCCACTTCGAGCACCGCGCGCTCTGCGCCGAGCGGAGC 9852
   :||| :||| :||| :||| :|||
147 LysGluLeuProArgGlnVal..... 153
   :||| :||| :||| :||| :|||
9853 GAGCTCTCTCCGCGCTCGACTCGCCACGACCAAGCCGCCGCGAG 9902
   :||| :||| :||| :||| :|||
154 .....PheGluGlnAspAlaSerGlyAspGluArgSerGlyProGluAsp 168
   :||| :||| :||| :||| :|||
9903 CACGCTTCGCGCGAGCGGAGACCAAGCGAGGTCGTCCTCTTTC 9952
   :||| :||| :||| :||| :|||
169 SerAsnAspSerArgArgGlySerGly.....SerProGluAspThrPr 183
   :||| :||| :||| :||| :|||
9953 CTGGCAAGGCTCCAGTGGAGGAGATGGCGCTCTCGCTCGACTCC 10002
   :||| :||| :||| :||| :|||
133 GArgSerSerGlyAlaSerSerIlePheAspLeuArgAsnValAla 200
   :||| :||| :||| :||| :|||
10003 TCGCGGCTCTCCGCGCTCAGC.....TCGACGATCGGCGCGC 10043
   :||| :||| :||| :||| :|||
200 spSerLeuLeu.....ProSerLeuLeuGluArg 209
   :||| :||| :||| :||| :|||
10044 GTCGCTCTCAGCTGAGTGAGCTGCTCGCGCTCGCTCGCGCGGAGC 10093
   :||| :||| :||| :||| :|||
210 AlaAlaProGluAspValAspArgArgAsnGluThrLeuArgAlaGlnH 226

```

[illegible]





273 GluIleuProIle.....Ph 278  
:::|::|  
10273 GCGATGCGCCCTGGCGAGAAAGCGCTCACACCGTCGGCGCAACGG 10322  
278 eGylIleuAlaIleuTyraAspValArgIuLysLysIleSerGluA 295  
||::|::|  
10323 CCGGATGGCGCGCTC..... 10338  
295 snPheTyraPheAspIleuAsnSerAspSerMetLysGlyLeuLeuArgAla 311  
:::|::|::|::|::|::|::|::|::|  
10339 .....GACCTGGCGCCCTCCGACTCCAGACTCCACTACCTGCTGCC 10377  
312 HisGlyThrHisProAlaIleSerThrLeuAla.....ArgSe 324  
|||  
10378 TGGGGCGACAGG.....CTCTCCACCGCGCGCGCTCAACAGCCCCAGGGC 10421  
324 rAlaIlePheSerValThrTyrrProSerProAspIlePheLeuValIleL 341  
:::|::|  
10422 TACCTGTATCCGCGAGCCCGCGCGCTGACGCGCGCTGCTC..... 10464  
341 yslLeuGluLysValLeuGluGlnGlyAspIleSerGluCysGluPro 357  
|||::|  
10465 .....GAGTCCTCAACCGCGCACCAAGGTG 10488  
358 TyrMetValLeuLysGluValAspThrAlaLysAsnLysGluLysLeuG 374  
:::|::|::|  
10489 TTCGCCCGCAAGATCCGCTGACCTACGCTCCACCTCCGCGCAAGAGGA 10538  
374 uLysLeuArg.....LeuAlaIleGlnGlnPheCysThrArgLeuGlu 389  
:|::|::|  
10539 CGCGGTCCAGACGACTCGCCGCA..... 10563  
389 rGlyrArgMetProPheAlaThrPheAlaValHisLeuAlaAsnIleVal 405  
|||::|::|::|  
10564 .....GCTCTAGCCAAACATGCT 10581  
406 SerSerAlaGlyGlnLeuAspArgAspSerAspSerGluGlnGluArg 422  
:::|::|  
10582 CCTCGGAGCTGCGAGCTCCCTTTATTTCGACCGTCACCGCGCACAGGCT 10631  
422 gProAlaThrPheAspArgArgArgGlyProGlnAspArgAlaSerS 439  
:::|::|  
10632 CGAC.....GGCTCCG 10642  
439 ergIysAspAlaCysSerPheSerGlyPheArgProAlaThrLeuThr 455  
:::|::|  
10643 AGCTGACGCGCGTACTGTATGAAACCTCCGCAACCGCTCTG... 10689  
456 ValThrAsnPhePheLysGlnGlnAlaGlnArgLeuSerAspGluAspLe 472  
|||::|  
10690 .....TTCCTGAGCGCGACGCGGCTCTCTCGAC...GATGG 10724  
472 uPheLysPheLeuAlaAspMetArg...ArgProSerSerLeuLeuArg. 487  
:::|::|  
10725 GCATCGCTTCCGTCGAGGTACAGCCCATCCGCTGTCAGGCTCGGCC 10774  
488 ArgLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAlaPr 504  
|||::|  
10775 TCCGGAGAACCTGAGCGCTACCGGCTGATCCGCTGCTGCGGCTCC 10824  
504 oGluAsnProHisPheCysLeuSerPro.....GluLeuLeuHisIle 518  
:::|::|  
10825 ATTGACGAGAAAGAGCCACCTGCGCGCTCTCTCTCTGCGCGGA 10874  
519 LysProTyrrProAspProArgGlyArg..... 527  
|||::|  
10875 GCTCTCTACCGAGAGCGCTGCGCTGACTGGAAGACTTCTTCGGGCCCT 10924  
528 .....ProThrLysGluIleLeuGluPheProAlaA 538  
|||  
10925 ACGCTCCCGCAAGAGTCTCTCCCTCCCGCAC.....CTACCCCTTCAGGGA 10968  
538 rgluValValThrAlaProHisThrSerTyrrArgAsnLeu..... 550

11069 GAGCGGTTCTGGCTGCGACTCTCCAGGACGAGACGCTTCCGAGCTGCCCT 11018  
551 LeuTyrrValTyrrProHisSerLeuAsnPheSerSerArgGln...GlySer 566  
:::|::|  
11019 CCGGAGGCTGACTGCTCGCGCCAGCAATCCCGCTGCTGCGCGCGCGCTCG 11068  
567 ValArgAsnLeuAlaValArgValGlnTyrrMetThrGlyGluAspProSe 583  
|||::|  
11069 CCTTCGCGACCGCGGTGGCTTCTCT...TTACAGGGCGGCTCTCCCTC 11115  
583 rGln.....AlaLeuPro..... 587  
|||  
11116 GCAGAGCACCCGTGGCTCGAAGGCCATCGCTTCGCGACACCATCTCT 11165  
588 .....ValIlePheGlyLysSerSerCysSerGluPheThrArgGlu 601  
|||::|  
11166 ACCGGGACCGGCTTCTGAGACTGCGCTGAGCTGCCACCGCGCTCG 11215  
602 AlaPheThrProValValTyrrHisAsnLysSerProGluPheTyrrGluG 618  
|||  
11216 GCCTCGACACCGTCTC.....GAAGA 11234  
618 uPheLysLeuHisLeuProAlaCysValThrGluAsnHisIleLeuAsp 635  
|||::|  
11235 GCTACAGCTCGAGGCGGCTCTGCTCTCCATCGCAGACGACGCTCTC 11283  
635 heThrPheTyrrHisValSerCysGlnProArgProGlyThrAlaLeuGlu 651  
|||::|  
11284 .....CTCAG 11289  
652 ThrProValGlyPheThrThrIleProLeuLeuGlnHisGlyArg..... 666  
|||  
11290 AACTCGTCGGG.....CCCGTGACGAGCGAGCAAGAGGC 11327  
667 .....LeuArgThrGlyProPhe. 672  
11328 GCCTCTTTCATAGCCGACAGAGAGAGCGGCTTCAGAGATGGCCCTCGGA 11377  
673 .....CysLeuProVal 676  
11378 CTCGCCACGACGCGCTCTCTCTCGCGCGACCCCATCTCTCCGCGC 11427  
677 SerValAspGlnProProProSerTyrrSerValLeuThrProAspValAl 693  
:::|::|  
11428 GATCTCACAGTGGCTCTCCCTCGAGTGCATC.....CCGGTGA 11468  
693 AlaLeuProGlyMetArgTrpValAspGlyHisLysGlyVal...PheSerV 709  
|||  
11469 CTCGGAAGGCTCTACGCAACCTCGCCCAACCTCGGGCTTGGCTACGGCC 11518  
709 aLysLeuThrAlaValSerSerValHisProGlnAspProTyrrLeuAsp 725  
|||  
11519 CCGAGTTCAGGGGCTCTCCGCTCTACAAAGCGCGCGACGAGCTCTTT 11568  
726 LysPhePheThrLeuValHisValLeuGluGlnGlyAlaPheProPheAr 742  
|||  
11569 GCCGAAGCCAGCTCCCGGAAGCGCGCGGAAGATGCCGCCGGTTTGC 11618  
742 gLeuLys.....AspThrValLeuSerGluGlyAsnValGlnG 755  
|||  
11619 CTCACACCTGCGCTGCTGACAGAGCGCTGATGCATGCGCTTGAAG 11668  
755 lngLeuAlaArgAlaSerLeuAla.....Ala 763  
|||  
11669 ACGAGCAGAGAGGAGCGTCTGCTCTGCTCTGCTGAGCGGAGTCTG 11718  
764 LeuArgLeuLaseProGluProLeuValAlaPheSerHisValle 780  
|||  
11719 CTGGGCTCGCTGCTGCGACACCTTGGCGGTGCTTCAC..... 11760  
780 uAspLysLeuValArgLeuValIleArgProProIleIleSerGlyGlnI 797  
|||::|

11761	.....CGTCCCAAGGCTGAATCCTCCGTCT	11785
797	LeValAsnleuGIyArgGIyAlaPheGluAlaMetAlaHisValSer	813
	:    :	:
11786	CGATGCTTCGGCGCGACGCCGCAGTGACCTTCCTCGGTGCAAGG	11835
814	LeuValHisArgSerIeuGluAlaInGln.....AspAlaArgGI	827
	:    :	:
11836	CTGGCATCGCGACGACGCTCCGCGCGACGCTCCGACC CGGCGACTTC	11885
827	yHisCysProGlnLeuAlaAlaTyrValHisTyrAlaPheArgLeuProg	844
	:	:
11886	CCACCATGATGCGCTC.....TTCCGCGTCGACT	11914
844	LyrHisLupProSerIeuProAspGIyAlaProProValThrValGlnAla	860
	:    :	:
11915	GGAAGCACAGCTCCAAACCCCACT...TCACCGCTGCGCGCCCGACGCGC	11961
861	AlaThrLeuAlaArgGIySer.....G1	868
	:	:
11962	GTCCTTCGCGGACAAGGGGCCACGATCTCGGCGCTCGACGCCCGCTCGC	12011
868	yArgProAlaSerIeuTyrLeuAlaArgSerIys...SerIleSerSers	884
	:    :	:
12012	CCGCTAAGCGCGACCTCGCTCCCTCGAAGCGCCCTCGACACAGGCGCTT	12061
884	eArgAsnProAspLeuAlaValAlaProGIySerValAspAspGIyValSer	900
	:    :	:
12062	C GGCTCCGGCGCTCGTCGTCGCGCCCC...TTCATGATCGACCGCGACGGC	12108
901	ArgG1LeuAlaSerIysLeuLeuHisGluGluLeuAlaLeuGlnTrpVa	917
	:    :	:
12109	GACCTCGTCCCGACG....GCCACGAG.....	12132
917	IValSerSerSerAlaValArgGluAlaIleGluGlnHisAlaTrpPheP	934
	:    :	:
12133	..GCCACCGGCTCGCACTCGCCTCTTGCA...GCTCGG...	12168
934	hePheGlnIeuMeValLysSerMetAlaLeuHisLeuLeuGlyGln	950
	:	:
12169	.....CTCGCGACGAGA	12180
951	ArgLeuAspThrProArgLysIeuArgPheProGlnArgPheLeuAspAs	967
	:    :	:
12181	CGCTCGCTCGCTCGCC.....	12198
967	PileThrAlaLeuValGIySerValGIyLeuGluValIleThrArgValH	984
	:    :	:
12199	.....CTCGCTCGCTCGCACCGACGCGCGCTCGCCACCCACACG	12238
984	IsLyAspValGIuLeuAlaGlnIleuAsnAlaSerLeuAlaPhePhe	1000
	:    :	:
12239	AAGACGACGCTCAG.....	12252
1001	LeuSerAspLeuLeuSerLeuValAspArgGIyPheValPheSerLeuVa	1017
	:	:
12253	.....GACCTCGC	12260
1017	LArgAlaHisTyrLysGlnValAlaThrArgLeuGlnSer...SerProAs	1033
	:    :	:
12261	TCACCGCGCGCTCGGGGCTCGCGCGCTCGCGCAAAGTAGACCCAG	12310
1033	nProAlaAlaLeuLeuThrIleuArgMetGlnPheThrArgILEueucyss	1050
	:    :	:
12311	ACCTCGCGCTTT.....C	12324
1050	eRHisGlnIisTyrValThrIleuAsnLeuProCys.....	1061
	:	:
12325	CTCGTCGACATCGACCTCGAGGAGCTCCACAGAGCCCTGCTAGAGCGC	12374
1062	.....CysProLeuSerProPro.....AlaSerProse	1071
	:	:
12375	GCTGACACAGAGACAGCCAGCTCGGCTCGGACCGGAAACCCCTTA	12424

```

1071  rProSerValSerSerThrThr_SerGlnSerSerThrPheSerSerGln 1087
12425  TCCGAGATGTGGCCACCAACGGCTCCAGGACGGCTCATCCGCCGCAA 12474A
1088  AlAProAspProIysValThrSerMetPheGluLeuSerGlyProPheAr 1104
12475  GCACCCACGAGGGCGCTC... 12492
1104  gGlnGlnHisPheLeuLeuLaGlyLeuLeuLeuThrGluLeuAlaLeuAl 1121
12493  ....CATATTCCGACCAAAAGCACCTTTGCAGCGGCTCGCCCTGTGG 12535B
1121  euGlnProGluAlaGluGlyAlaPheLeuLeuHisLysLysAlaLeuSer 1137
12536  ACCGCCCCGAGGGCCAG..GGCCCCCTCGCACCAGCGCCAAATGCCCATC 12582B
1138  AlAlaValHisSerLeu..... 1142
12583  GCGCTGCACGCGCAGAGGGCTCACTTCGCGATGTGTCGCACACCTTTGG 12632B
1143  .....LeuGlySerGlyHis..... 1146
12633  CATGTAATCCGGGGAGGAGCGCGCGCGCTTCGGAGGCGCAAGCGCGGCATCG 12682B
1147  ....AspThrAspProArgTyrAlaGluAlaThrValLysAlaArgVal 1161
12683  TTACTAGATGTCGGTCCAGAGTGTCTCCCGATACACGTAAGCAGCGGGTGG 12732B
1162  .....AlaGlu 1163
12733  ATGGGGGCTCTGGCGGACGCTTTGGTCTCCACGGCCATCGCGCAGCGCGG 12782B
1163  uLeuTyrLeuProLeuLeuSerTleAlaArgAspThrLeuProArgLeuH 1180
12783  CAAAGATCTGCGCC...ATCCCCACGCGCT 12808
1180  IsAspPheAlaGluGly_ProGlyGlnArgSerArgLeuAla...SerMe 1195
12809  GGCCTTTCGCCCAAGCGCCGACAGCTGCCATCATCTATCTCACGCCCTTAC 12858
1195  tLeuAspSerAspThrGluGlyLeuGlyLysPheAlaGluThrTleAspP 1212
12859  TATGGACCTGTGATCTCGGCG...ATCTGAACCAACCAATCAACG 12899
1212  roSerValAlaMet.....AlaIleAlaGlyGlyProLeuAlaPro 1225
12900  TGTCCCTCATCTCATGGCGCGCGCGCGGGCGGTGGGAGCGCGCGCTTACG 12949
1226  GlySerArgAlaSerTleSerGlnGlyProProThrAlaSerArgAlaGlu 1242
12950  TCGCACGCCACCTCGCGCGCGGAGGTCTTGGCACACGCCACATGCAGGAG 12999
1242  yCysAlaLeuSerAlaGluSerSerArgThrLeuLeuAlaCys..... 1256
13000  TGGAGGCGCTCTCGCGCGGTCTGCAGTGCACATGCAGCTCGCGCTCTC 13049
1257  ..ValLeuThrValLeuLysAsnThrGlu.....ProAlaLeuGln 1270
13050  ACGTGAACCTGGGCTTCGAGACAGCATCTCTCGCTCCACGATGGGCCGG 13099
1271  ArgTyrPheAlaThrAspLeuThrLeu 1278
13100  GCATGGATGTCTCTCTCGCATGTCT 13123
seq_name: /cgm2_6/ptodata/2/lna/5A_COMB_seq:US-08-457-335A-6
seq_documentation_block:
; Sequence 6, Application US/08457335A
; Patent No. 5723759
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas M.
; APPLICANT: Ligon, James M.

```





```

10378 TGGGGCGACAG.....CTCTCCACCGCGCGCTCAACAGCCCGACGGC 10421
324  rAlaIlePheSerValThrTyProSerProAspIlePheLeuValIleL 341
      ::::: ||| ::::: ||| ::::: |||
10422 TACCTGTATCCGGCGAGCCCGCGCGCTGTC..... 10464
341  yslLeuGluLysValLeuGlnGlnIleAspIleSerGluCysGluPro 357
      ||::: ||::: ||::: ||:::
10465 .....GACGTCCTCACCGCGCCACCAAGGTG 10488
358  TyrMetValLeuLysGluValAspThrAlaLysAsnLysGluLysLeuG1 374
      ::::: ||::: ||::: ||::: ||::: ||:::
10489 TTCCGCCGCAAGATCCGCTGCTACGCTCCACCTCCGCGCACATGGA 10538
374  uLysLeuArg.....LeuAlaIleGlnIlePheCysThrArgLeuGly 389
      ::::: ||::: ||::: ||::: ||::: ||:::
10539 CGCGGTCCAGAGAGAGCTCCCGCA..... 10563
389  rGlyThrArgMetProPheAlaThrPThrAlaValHisLeuAlaAsnIleVal 405
      ||::: ||::: ||::: ||::: ||::: ||:::
10564 .....GCTCAGCCACATCGCT 10581
406  SerSerAlaGluGlnLeuAspArgAspSerAspSerGluGluArg 422
      ::::: ||::: ||::: ||::: ||::: ||:::
10582 CCTCGGACGTGGAGCTCCTTTATTCGACCGTCACCGCGCACCGCT 10631
422  gProAlaThrPThrAspArgArgArgGlyProGlnAspArgAlaSer 439
10632 CGAC.....GGCTCCG 10642
439  eArgLysAspArgAlaCysSerPheSerGlyPheArgProAlaThrLeuThr 455
      ::::: ||::: ||::: ||::: ||::: ||:::
10643 AGCTCGAGCGGCGTACTGTATCGAAACCTCGGCAACCGCTCTG... 10689
456  ValThrAsnPhePheLysGlnGlnIleGluArgLeuSerAspLysLeu 472
      ||::: ||::: ||::: ||::: ||::: ||:::
10690 .....TTCTCGAGCGCGAGCGAGCGCTCTCTGAC...SATGG 10724
472  uPheLysPheLeuAlaAspMetArg...ArgProSerSerLeuLeuArg 487
      ::::: ||::: ||::: ||::: ||::: ||:::
10725 GCATCGCTTCTCCGTGAGGTGAGCCCGCCATCCGCTCAGCGCGCC 10774
488  ArgLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAlaThr 504
      ||::: ||::: ||::: ||::: ||::: ||:::
10775 TCGCGCAGACCTGCGAGCGCTCACCGCTCGATCCGCTCGTGGCTCC 10824
504  oGluAsnProHisPheCysLeuSerPro.....GluLeuLeuHisIle 518
      ||::: ||::: ||::: ||::: ||::: ||:::
10825 ATTCGACGAGAGAGAGCCACCTCGCGCGCTGCTCTCTGGGCGGA 10874
519  LysProTyProAspProArgGlyArg..... 527
10875 GCTCTCTACCCGAGCGCTCGCTCGATCGAAGGACTTCTTGGCGCCT 10924
528 .....ProThrLysGluIleLeuGlnPheProAla 538
10925 AGCGTCCCGCAAGGTCTCCCTCCAC.....CTACCCCTTCCAGCGA 10968
538  rGluValIleValAlaProHisThrSerTyArgAsnLeu..... 550
10969 GAGCGGTCTGCTCGACGTCTCCAGGACGAAGCTTCCGAGCTGGCT 11018
551  LeuTyValTyProHisSerLeuAsnPheSerSerArgGln...GlySer 566
      ::::: ||::: ||::: ||::: ||::: ||:::
11019 CCGCAGCGCTGAGCTCGCGCAGCAATCCGCTGCTCGCGCGCGCGCT 11068
567  ValArgAsnLeuAlaValArgValGlnTyMetThrGlyLysAspPro 583
      ||::: ||::: ||::: ||::: ||::: ||:::
11069 CCTCGCCGACCGCGGTGCTTCTCT...TTACAGGGGCGCTCTCCCTC 11115
583  rGln.....AlaLeuPro..... 587
      ||::: ||::: ||::: ||::: ||::: ||:::
11116 GCAGAGACACCGCTGGCTCGAAGCCATGCGCTCTGGACACCCGACT 11165

```

```

588 .....ValIlePheGlyLysSerSerCysSerGluPheThrArgGlu 601
      ||::: ||::: ||::: ||::: ||::: ||:::
11166 ACCGGGACACGGCTTCTGAGCTGCGCTCGACGCTCGCCGCGCTCG 11215
602  AlaPheThrProValValTyThrHisAsnLysSerProGlnPheTyGlnG1 618
      ||::: ||::: ||::: ||::: ||::: ||:::
11216 GCCTCGACACCGCTC.....GAAGA 11234
618  uPheLysLeuHisLeuProAlaCysValThrGluAsnHisLeuLeuP 635
      ||::: ||::: ||::: ||::: ||::: ||:::
11235 GCTCAGCTCGAGCGCCCTCTGCTCCCATCGCAGAGCACCGCTCTC 11283
635  heThrPheTyThrHisValSerCysGlnProArgProGlyThrAlaLeuGlu 651
      ||::: ||::: ||::: ||::: ||::: ||:::
11284 .....CTCAG 11289
652  ThrProValGlyPheThrTrIleProLeuGlnHisGlyArg..... 666
      ||::: ||::: ||::: ||::: ||::: ||:::
11290 ATCTCCGTGGG.....CCGCTGAGCAGCGAGCAGCAAGGGC 11327
667 .....LeuArgThrGlyProPhe 672
11328 GCTCTTTTCCATAGCCGACAAGAGAGCGCTTCAGATGCGCCCTGA 11377
673 .....CysLeuProVal 676
11378 CTCGCCACCGCAGCGGCTCTCTCGCGCGGACCCCATCCCTCGCCG 11427
677  SerValAspGlnProProSerTySerValLeuThrProAspValAl 693
      ::::: ||::: ||::: ||::: ||::: ||:::
11428 GATCTCCAGAGTGCGCTCCCTCGAGTGCATC.....CGGATGA 11468
693  aLeuProGluMetArgTrpValAspGlyHisLysGlyVal...PheSerV 709
      ||::: ||::: ||::: ||::: ||::: ||:::
11469 CTTGAAAGGCTCTACGCAACCTCGCCACCTCGGCTTCCCTACGAGC 11518
709  aGluLeuThrAlaValSerSerValHisProGlnAspProTyLeuAsp 725
      ||::: ||::: ||::: ||::: ||::: ||:::
11519 CGAGTTCCAGAGGCTCGCTCCGCTCAAGCCGCGGAGACGACTCTT 11568
726  LysPhePheThrLeuValHisValLeuGluGluLysAlaPheProPhe 742
      ||::: ||::: ||::: ||::: ||::: ||:::
11569 GCCGAAGCAAGCTCCGGAAGCGCGGAAAAGATGCGCGCGGTTC 11618
742  gLeuLys.....AspThrValLeuSerGluGlyAsnValGlu 755
      ||::: ||::: ||::: ||::: ||::: ||:::
11619 CTTCCACCTTGCCTGCTCGACAGCGCTGCATGCACTGGCTTTGAG 11668
755  IngLuLeuArgAlaSerLeuAla.....Ala 763
      ||::: ||::: ||::: ||::: ||::: ||:::
11669 ACGAGCAGAGAGGAGAGGTGCTCTGCGCTTCTGTGAGGAGGTCTG 11718
764  LeuArgLeuAlaSerProGlnProLeuValAlaPheSerHisValLe 780
      ||::: ||::: ||::: ||::: ||::: ||:::
11719 CTGCGCTCTCGTGGTGCACCACTTGCCTGCGCTTCCAC..... 11760
780  uAspGlyLeuValArgLeuValIleArgProProIleIleSerGlyLys 797
      ||::: ||::: ||::: ||::: ||::: ||:::
11761 .....GCTCCCAAGGTGAATCTCTCTCT 11785
797  LeValAsnLeuGlyArgGlyAlaPheGluAlaMetAlaHisValSer 813
      ::::: ||::: ||::: ||::: ||::: ||:::
11786 CGATGCTGCTGGCGCAGCGGAGGTGATACCTCTGTGCTGGTGCAAGC 11835
814  LeuValHisArgSerLeuGluAlaIleGln.....AspAlaArgG1 827
      ||::: ||::: ||::: ||::: ||::: ||:::
11836 CTCGCATGGGAGAGAGCTCGCGCGCAGCTCCGACCCCGCACATTC 11885
827  yHisCysProGlnLeuAlaIleValIleTyThrAlaPheArgLeuPro 844
      ||::: ||::: ||::: ||::: ||::: ||:::
11886 CCACCATGATGCGCTC.....TTCCGCGTGA 11914

```

```
844 1yThrgluProSerLeuProAspGlyAlaProProValThrValGlnAla 860
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
11915 GAGAGGAGAGTCCAAAGCCCCACT..TCACCGCTGCGCGCCGAGCGAGC 11961
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
861 AlaThrLeuAlaArgLysSer.....G1 868
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
11962 GTCCTTCCTCGGCA?AGGCGCGACAGATCTCGGCTCGAGACCCCGCTCGC 12011
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
868 yArgProAlaSerLeuTyThrLeuAlaArgSerLys...SerLysSerLys 884
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12012 CCGCTACGCGCGAGCTCGCTCGCTCGGAGCGCCCTCGACACAGGGGCGCTT 12061
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
884 eAspProAspLeuAlaValAlaProGlySerValAspAspGlyValSer 900
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12062 CCGCTCGCGGCTCGCTCGCGCCCC..TTCATGATCGACGCGGAGGC 12108
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
901 ArgLLeuAlaSerLysLeuLeuHisGluLeuAlaLeuGlnTrpVal 917
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12109 GACCTCGTCCGAGC.....GCCACGAG..... 12132
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
917 yValSerSerValAlaValArgGluAlaLeuGlnHisAlaTrpPhe 934
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12133 .....GCCACCGCGCTCGCGACTCGCCCTCTGCA...GCCG..... 12168
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
934 hePheGlnLeuMetValLysSerMetAlaLeuHisLeuLeuGlyGln 950
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12169 .....CTCGCGCGAGCA 12180
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
951 ArgLeuAspThrProArgLysLeuArgPheProGlyArgPheLeuAspAs 967
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12181 CGCCTCGCTCGTCCGCG..... 12198
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
967 PileThrAlaLeuValGlySerValGlyLeuGluValLeuThrArgValH 984
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12199 .....CTCGCTCGTCCGCGCGCGCGCGCTCGCGCACACACACCG 12238
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
984 iLysAspValGluLeuAlaGlnHisLeuAsnAlaSerLeuAlaPhePhe 1000
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12239 AAGACGACGCTCAAG..... 12252
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1001 LeuSerAspLeuLeuSerLeuValAspArgGlyPheValPheSerLeuVal 1017
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12253 .....GACCTCGC 12260
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1017 iArgAlaHisTyrlYsGlnValAlaThrArgLeuGlnSer...SerProAs 1033
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12261 TCACGCGCGCTCGGCGGCTCGCGCTCGCGCAAAAGTAGACACCCAG 12310
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1033 nProAlaAlaLeuLeuThrLeuArgMetGlnPheThrArgLeuLeuGlyS 1050
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12311 ACCTCCGCTCTT.....C 12324
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1050 eHisGlnHisTyrlValThrLeuAsnLeuProGlyS..... 1061
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12325 CTGCTGACATCGACCTCGACGAGGCTCCGACGAGGCTCGTAGAGCGC 12374
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1062 .....CysProLeuSerProPro.....AlaSerProse 1071
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12375 GCTCGACACAGAGAAAGCGCACTCGCCCTCGCAACGGAACCCCTCA 12424
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1071 rProSerValSerSerThrThr.SerGlnSerSerThrPheSerSerGln 1087
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12425 TCCTCGAGGTTGGCGCAACGACCTCGACGAGCGCTCATCTCGCGCGCA 12474
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1088 AlaProAspProLysValThrSerMetPheGluLeuSerGlyProPheArg 1104
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12475 GCAACCGACGTCGCGCTC..... 12492
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1104 gGlnGlnHisPheLeuAlaGlyLeuLeuLeuThrGluLeuAlaLeuAlaL 1121
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12493 .....CATATCTCGACCAAGGACACTTCGACGCGCTCGCGCTCGCG 12535
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1121 eugluProGluAlaGlnGluAlaPheLeuLeuHisLysValAlaLeuSer 1137
```

```
12536 ACCCCCCGAGGCGCAG...GGCGCCCTCGCACACAGGCAAGTCCGATC 12582
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1138 AlaValHisSerLeu..... 1142
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12583 GCCGTCACAGCGGAGGCTCACTTCGCGATGTCGTGACACCTTTGG 12632
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1143 .....LeuGlyHis..... 1146
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12633 CATGATCCGGGCGAGCGCGCGCTCGAGAGCGAGAGCGGCGCATCG 12682
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1147 .....AspThrAspProArgTyrlAlaGlnAlaThrValLysAlaArgVal 1161
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12683 TTACTGAGTGCCTCCAGATGTCCTCCGATACCGCTAGGCGAGCGGATG 12732
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1162 .....AlaG1 1163
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12733 ATGGGGGCTTCGGCGCAGCCTTTGTTGCCACGGCCATCGCCGAGCGCG 12782
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1163 uLeuTyrlLeuProLeuLeuSerLLeuAlaArgAspThrLeuProArgLeuH 1180
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12783 CATGATCTGCCCC.....ATCCCGACGCGCT 12808
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1180 iAspPheAlaGluGly.ProGlyGlnArgSerArgLeuAla...SerLe 1195
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12809 GGTCTTCGCCACAGCGCGCACGCTCCCATCATCTATCTACCGCTTAC 12858
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1195 tLeuAspSerAspThrGlnGlyGluGlyAspLLeuAlaGlyThrLLeuAsn 1212
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12859 TATGACCTGCTGATCTCGGC.....ATCGAACAACCATTCACAG 12899
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1212 roSerValAlaMet.....AlaLLeuAlaGlyGlyProLeuAlaPro 1225
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12900 TGAGCTCATCATAGCGCGCGCGCGCGCTCGGAGCGCGCGCTCGAGC 12949
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1226 GlySerArgAlaSerLLeuSerGlnGlyProProThrAlaSerArgAlaG1 1242
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12950 TCACAGCGCACCTCGCGCGCGAGGCTTTGGCCACCGCGCAGTCCAGGAG 12999
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1242 yCysAlaLeuSerAlaGlnSerSerArgTyrlLeuLeuAlaCys..... 1256
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
13000 TGAGCGCTCTCCGCGCGCTCGCTTCGACAGTCCGACCTCGCGCTCTC 13049
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1257 ..ValLeuTyrlValLeuLysAsnThrGln.....ProAlaLeuLeuGln 1270
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
13050 ACCTGACCTGGGCTTCGAGACAGCACTTCTCGCGCTCGACGATGGCGCG 13099
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1271 ArgTyrlAlaThrAspLeuThrLeu 1278
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
13100 GCATGATGTGCTCCGACCTGTC 13123
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:us-08-729-214-6
seq_documentation_block:
; Sequence 6, Application US/08729214
; Patent No. 5817502
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Hammer, Phillip E.
; APPLICANT: van Pee, Karl-Heinz
; APPLICANT: Kirner, Sabine
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 520 White Plains Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
```



```

406 SerSerAlaGlyLeuAspArgSerSerGluGlyLysArg 422
    ::::: ||| ||| |||
10582 CCGGAGCGCTCCCTTTATTCAGCCGACCGCGCGCT 10631
    ::::: ||| ||| |||
422 gProAlaTrpThrAspArgArgArgGlyProGlnAspArgAlaSer 439
    ::::: ||| ||| |||
10632 CGAC.....GGCTCCG 10642
439 erGlyAspArgAlaCysSerPheSerGlyPheArgProAlaThrLeuThr 455
    ::::: ||| ||| |||
10643 AGCTCGACGCGCGTACTGATCGAAACCTCGGCAACCGCTCTG... 10689
456 ValThrAsnPhenPheGlyGlnGlyAlaArgLeuSerAspLysPle 472
    ::::: ||| ||| |||
10690 .....TTCGAGCGCGACCGCGCTCTCTCGAC...GATGG 10724
472 uphGlyPheLeuAlaAspMetArg...ArgProSerSerLeuLeuArg 487
    ::::: ||| ||| |||
10725 GCATCGCTTCGTCGAGGTGAGCCGCCATCCCGTCTCAGCTGCCCC 10774
488 ArgLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAlaP 504
    ::::: ||| ||| |||
10775 TCCGCGAGACCTCGAGCGCTCACCCTGATCCCGTCTCGTCTCGCTCC 10824
504 ogLysAsnProHisPheCysLeuSerPro.....GluLeuLeuHisIle 518
    ::::: ||| ||| |||
10825 ATTCGACGAGAGAGAGAGCCACTCGCCCGCTGCTCTCTCTCGGCGGA 10874
519 LysProLysProAspProArgGlyArg..... 527
    ::::: ||| ||| |||
10875 GCTCTCTACCCGAGCGCTCGCGCTGAGTGAAGACTCTTCTGCGGCCCT 10924
528 .....ProThrLysGluLeuLeuGlnPheProAla 538
    ::::: ||| ||| |||
10925 ACCGCTCCCGCAAGCTCTCCCTCCAC.....CTACCCCTTCCAGCGA 10968
538 rglValValTyrAlaProHisThrSerTyrArgAsnLeu..... 550
    ::::: ||| ||| |||
10969 GAGCGGTCTGCTGCGAGCTCTCCAGGACGAGACGCTTCCGAGCTGCGCT 11018
551 LeuTyrValTyrProHisSerLeuAsnPheserSerArgLys...GlySer 566
    ::::: ||| ||| |||
11019 CCGCAGGCTGACCTCGCGCAGACCAATCCCGCTCGCGCGCGCGCTG 11068
567 ValArgAsnLeuAlaValArgValGlnTyrMetThrGlyGluAspPro 583
    ::::: ||| ||| |||
11069 CCGTCCCGCAGCGCGCTGCTCTCT...TTACAGGCGGCTCTCTCTC 11115
583 rglN.....AlaLeuPro..... 587
    ::::: ||| ||| |||
11116 GCAGAGACCCGTCGTCGAGGCGCATGCGTCTCGGACACCATCTCT 11165
588 .....ValIlePheGlyLysSerSerCysSerGluPheThrArgLys 601
    ::::: ||| ||| |||
11166 ACCGGGACCGGCTTCTGAGCTGCGCCCTCGACGCGCCACCGCGTGC 11215
602 AlaPheThrProValValTyrHisAsnLysSerProGlnPheTyrGlnI 618
    ::::: ||| ||| |||
11216 GGCTGACACCGCTC.....GAGA 11234
618 uphGlyLeuHisLeuProAlaCysValThrGlnAsnHisLeuLeuP 635
    ::::: ||| ||| |||
11235 GCTCAGGCTCGAGGCGCTCTGCTCCATCGCAGGACGACCGTCTCTC 11283
635 heThrPheTyrHisValSerCysGlnProArgProGlyThrAlaLeuGln 651
    ::::: ||| ||| |||
11284 .....CTCCAG 11289
652 ThrProValGlyPheThrTrpIleProLeuLeuGlnHisGlyArg..... 666
    ::::: ||| ||| |||
11290 ATCTCGCTGGG.....CCCGTGGACGACGACGACGACGACGACG 11327
667 .....LeuArgThrGlyProPhe. 672

```

---

```

11328 GCTCTTTTCATAGCCGACAAGAGAGCGGCTTCAGATGACCCCTGGA 11377
    ::::: ||| ||| |||
673 .....CysLeuProVal 676
    ::::: ||| ||| |||
11378 CTCGCCAGCGCAGCGGCTCTCTCGCGCGGACCCCATCTCTCTCCGCG 11427
677 SerValAspGlnProProSerTyrSerValLeuThrProAspValAl 693
    ::::: ||| ||| |||
11428 GATCTCCAGATGCGCTCCCTCGAGTGCATC.....CCGGTGA 11468
693 aLeuProGlyMetLysTyrValAspGlyHisLysGlyVal...PheSerV 709
    ::::: ||| ||| |||
11469 CCTCGAAGGCTCTACGCAACCTCGCAACCTCGGCTCTCTCTCTCTCT 11518
709 aGluLeuThrAlaValSerSerValHisProGlnAspProTyrLeuAsp 725
    ::::: ||| ||| |||
11519 CCGAGTTCAGGCGCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11568
726 LysPhePheThrLeuValHisValLeuGlnGlyAlaPheProPheAr 742
    ::::: ||| ||| |||
11569 GCCGAAGCCAGCTCCGGAAGCGCGGCAAAAGATGCCGCGGTTTGC 11618
742 gLeuLys.....AspThrValLeuSerGluGlyAsnValGlyG 755
    ::::: ||| ||| |||
11619 CTTCCACCTCGCGCTCTGACAGCGCTCTGACACTGCGCTTGTGAGG 11668
755 GlnGluLeuArgAlaSerLeuAla.....Ala 763
    ::::: ||| ||| |||
11669 ACGACAGAGAGAGAGCGCTCTGCGCTCTCTCTCTCTCTCTCTCTCT 11718
764 LeuArgLeuAlaSerProGluProLeuValAlaPheSerHisValLe 780
    ::::: ||| ||| |||
11719 CTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11760
780 waSprLysLeuValArgLeuValIleArgProProIleIleSerGlyGln 797
    ::::: ||| ||| |||
11761 .....CTGCCAAGGCTGAACTCTCTCTCT 11785
797 leValAsnLeuGlyArgGlyAlaPheGlnAlaMetAlaHisValValSer 813
    ::::: ||| ||| |||
11786 CGATGCTCTCTGCGCGAGCGCGCAGGTGACCTCTTGGCTGCTCGCAAGCG 11835
814 LeuValHisArgSerLeuGluAlaAlaGln.....AspAlaArgGly 827
    ::::: ||| ||| |||
11836 CTGCGCATGCGGACGAGCTCCGCGCAGCTCCGACCCCGGAGCTCTC 11885
827 yHisCysProGlnLeuAlaIleTyrValHisTyrAlaPheArgProG 844
    ::::: ||| ||| |||
11886 CCACCATGATCGCTC.....TTCGCGCTCACT 11914
844 LyrThrGluProSerLeuProAspGlyAlaProProValThrValAla 860
    ::::: ||| ||| |||
11915 GCAGCAGACTCCAAAGCCCACT...TCACGCGCTCGCGCGCGAGGCG 11961
861 AlaThrLeuAlaArgGlySer.....G 868
    ::::: ||| ||| |||
11962 GTCTCTCTGCGACAGCGCGCAGATCTCGGCTCGACCCCGCGCTCGC 12011
868 YArgProAlaSerLeuTyrLeuAlaArgSerLys...SerIleSerSers 884
    ::::: ||| ||| |||
12012 CCGCTAGCGCGACCTCTGCGCTCCGAAAGCGCTCGGACGAGGCGCTT 12061
884 erAsnProAspLeuAlaValAlaProGlySerValAspArgGlyValSer 900
    ::::: ||| ||| |||
12062 CGCTCCCGGCTCTGCTGCGCGCC...TTCACTGATCGACGCGGAGCG 12108
901 ArgGluLeuAlaSerLysLeuLeuHisGluGluLeuAlaLeuGlnTrpVa 917
    ::::: ||| ||| |||
12109 GACTCTGCTCGGAGC.....GCCAGAG..... 12132
917 lValSerSerSerAlaValArgGluAlaIleLeuGlnHisAlaTrpPhe 934
    ::::: ||| ||| |||

```

```

12133 .....GCCACCGCGCTCGACCTGCGCCCTCTTGCAA...GCCTGG.... 12168
934 hepheGlnleuMetValLysSerMetAlaLeuHisLeuLeuGln 950
12169 .....CTCGCGGACGAA 12180
951 ArgLeuAspThrProArgLysLeuArgPheProGlyArgPheLeuAspAs 967
12181 CGCCTCGCCTCGCGCC..... 12198
967 PileThrAlaLeuValGlySerValGlyLeuGlnValIleThrArgValH 984
12199 .....CTGTCCTCTGCACCGCGACGCGCTCGCCACCGCACCG 12238
984 IsuLysAspValGluLeuAlaGlnHisLeuAsnAlaSerLeuAlaPhe 1000
12239 AAGACACGCTCAAG..... 12252
1001 LeuSerAspLeuLeuSerLeuValAspArgGlyPheValPheSerLeuVa 1017
12253 .....GACCTCGC 12260
1017 IArgAlaHisThrLysGlnValAlaIleThrArgLeuGlnSer..SerProAs 1033
12261 TCACGCGCGCGCTCTGGGGGCTCGCGCGCTCCGCGCAAGTGACGACCGCAG 12310
1033 nProAlaAlaLeuLeuThrLeuArgMetGluPheThrArgGlyLeuGlyss 1050
12311 ACCTCCGCTCTT.....C 12324
1050 exHisGlnHisThrValThrLeuAsnLeuProGys..... 1061
12325 CTGCTGACATGACACTCGACGCGGCTCGCGACGCGCTGTAAGCGC 12374
1062 .....CysProLeuSerProPro.....AlaSerProse 1071
12375 GCTCGACACAGAGAAAGCCACTCGCTCGCGCAAGCGGAAACCCCTCA 12424
1071 rProSerValSerSerThrThr..SerGlnSerSerThrPheSerGln 1087
12425 TCCCGAGGTTGGCGCAACCAAGCTCGACGAGCGCTCATCCCGCGCA 12474
1088 AlaProAspProLysValThrSerMetPheGlnLeuSerGlyProPheAr 1104
12475 GCACCCACGAGGCGCTC..... 12492
1104 gGlnGlnHisPheLeuAlaGlyLeuLeuThrGluLeuAlaLeuAla 1121
12493 .....CATATTCCGACCAAGGACCTTCGACGCGCTCGCTCTGTG 12535
1121 euGluProGluAlaGlnGlyAlaPheLeuLeuHisLysLysAlaIleSer 1137
12536 ACCGCCCCGAGGCCAG...GCCGCTCGACACGCGCAAGTCCGCAATC 12582
1138 AlaValHisSerLeu..... 1142
12583 GCGCTGACGCGGAGGCTCAACTCCGCGATGTCGTGACACCTTGG 12632
1143 .....LeuGlyGlnHis..... 1146
12633 CATGTATCGGCGGACGCGCGCGCTCGAGGCGAAGCGCGGCGCATG 12682
1147 .....AspThrAspProArgLysValAlaIleThrValLysAlaArgVal 1161
12683 TTACTGAAGTCGTCGAGGTCTCCCGATACACCTAGCGACGCGGTG 12732
1162 .....AlaGln 1163
12733 ATGGGGGCTTTGGCGACGCTTGGTCCACGCGCATGCGCGACGCGCG 12782
1163 uLeuThrLeuProLeuLeuSerIleAlaArgAspThrLeuProArgLeuH 1180
12783 CATGATCTCGCC.....ATGCCCAAGCGCT 12808

```

```

1180 IsaPheAlaGlnGly..ProGlnArgSerArgLeuAla...SerHe 1195
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
12809 GGTCTTCGCCCAAGCGCGCGCTGCCCATCATCATGACCGGCTAC 12858
1195 LLeuAspSerAspThrGlnGlyGlnLysAspIleAlaGlyThrIleAsn 1212
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
12859 TATGACTCTCGATCTCGGGC.....ATCTGAACCAATCATCAG 12899
1212 roSerValAlaMet.....AlaIleAlaGlnGlyProLeuAlaPro 1225
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
12900 TGTCCTATCATCGGCGCGCGCGCGCTGGGACGCGCGCTTCAGC 12949
1226 GlySerArgAlaSerIleSerGlnGlyProProThrAlaSerArgAla 1242
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
12950 TCGACGCCACACTCGGCGCGAGGTCTTGCCACCGCACTCCAGGAG 12999
1242 yCysAlaLeuSerAlaGlnSerSerArgThrLeuLeuAlaCys..... 1256
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13000 TGGAGCGCTCTCGCGCGCTCGGCTTCGACGATGCGCATGCGCTC 13049
1257 ..ValLeuThrValLeuLysAsnThrGlu.....ProAlaLeuGln 1270
13050 ACGTGACCTGGGCTTCGAGCAGCACTTCTGCGCTCCACGATGGCGCG 13099
1271 ArgThrAlaThrAspLeuThrLeu 1278
|||:|||||:|||||:|||||:|||||:|||||:|||||:
13100 GCATGATGTCGCTCGACTGTC 13123

seq_name: /cgn2_6/ptodata/2/lna/6A.COMB.seq:US-09-028-934-6

seq_documentation_block:
; Sequence 6, Application US/09028934
; Patent No. 6117670
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Lam, Steven T.
; APPLICANT: Hammer, Philip E.
; APPLICANT: van Pee, Karl-Heinz
; APPLICANT: Kirner, Sabine
; APPLICANT: Young, Thomas R.
; TITLE OF INVENTION: Pyroclitric Biosynthesis Genes and Uses
; TITLE OF INVENTION: Theoret
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6117670artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/028,934
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1506/CIP7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587

```







```

12239 AAGACGACGTCACG..... 12252
1001 LeuSerAspLeuLeuSerLeuValAspArgGlyPheValPheSerLeuVal 1017
12253 .....GACCTCGC 12260
1017 LArgAlaHisTyrLeuGlnValAlaThrArgLeuGlnSer..SerProAs 1033
12261 TCACGCGCCGCTCTGGGGCTCGCGCGCTCCGCGCAAGTGCACACCCAG 12310
1033 nPrAlaAlaLeuLeuThrLeuArgMetGluPheThrArgIleLeuGlyss 1050
12311 ACCTCCGCGCTCT.....C 12324
1050 eRhHisGluHisTyrValThrLeuAsnLeuProGlyss..... 1061
12325 CTCGTGCACATCGACCTCAGGAGGCGCTCCACAGGCGCCCTGCTAGGCGC 12374
1062 .....CysProLeuSerProPro.....AlaSerProse 1071
12375 GCTCGACACAGAGACGACGCGCTCTCCGACAGGAGAAACCCCTCA 12424
1071 rProSerValSerSerThrThr..SerGlnSerSerThrPheSerSerGln 1087
12425 TCCTCGAGGTGGCCCAACACGCTCGACGAGCGGCTCATCCGCGCCAA 12474
1088 AlaProAspProLysValThrSerMetPheGluLeuSerGlyPropheAr 1104
12475 GCACCCACGTCGGCGCTC..... 12492
1104 gGlnGlnHisPheLeuAlaGlyLeuLeuLeuThrGluLeuAlaLeuAla 1121
12493 .....CATATTCGACCAAGGACCTTCGACGCGCTCGCTCGCTC 12535
1121 euGluProGluAlaGluGlyAlaPheLeuLeuHisLysLysAlaIleSer 1137
12536 ACGCCCCCGAAGCCACG...GCGCCCTCGACACGCGCAAGTCCGCATC 12582
1138 AlaValHisSerLeu..... 1142
12583 GCCGTGCACGCGGACAGGGCTCACTTCGCGATGTCGTGCACACCTTGG 12632
1143 .....LeuGlyGlyHis..... 1146
12633 CATGTATCGGGCGACGCGCGCTCGAGGCAAGCGCGGCATCG 12682
1147 .....AspThrAspProArgTyrAlaGluAlaThrValLysAlaArgVal 1161
12683 TTACTGAAGTGGTCCAGGTGTCTCCGATACACCGTAGCGGACCGGTG 12732
1162 .....AlaGlu 1163
12733 ATGGGGGTCTTCGGCGCAGCCTTGGTCCACGCGCATCGCGGAGCGCG 12782
1163 uLeuTyrLeuProLeuLeuSerIleAlaArgAspThrLeuProArgLeuH 1180
12783 CATGATCTGCCCC.....ATCCCCACGCGCT 12808
1180 IsAspPheAlaGluGly..ProGlyGlnArgSerArgLeuAla...SerMe 1195
12809 GGTCTTCGCCCAAGCCGCGCGCTCCCATCTATCTACCGCGCTAC 12858
1195 tLeuAspSerAspThrGluGlyGlyAspIleAlaGlyThrIleAsn 1212
12859 TATGACTCTGCTCATCTCGGC.....ATCTGAACCAACCAATCAACG 12899
1212 roSerValAlaMet.....AlaIleAlaGlyGlyProLeuAlaPro 1225
12900 TGTCTCTACATGCGCGCGCGCGGTGGGCGCGCGCGCTTACGC 12949
1226 GlySerArgAlaSerIleSerGlnGlyProProThrAlaSerArgAlaGlu 1242

```

```

12950 TCGCACGCCACTCGGCGCGCGAGTCTTTGCCACCGCCAGTCCAGGAAG 12999
1242 yCysAlaLeuSerAlaGluSerSerArgThrLeuLeuAlaCys..... 1256
13000 TGGAGCGCTCTCCGCGCGCTCGCTTCGACGATGGCGACCTCGCGCTC 13049
1257 ..ValLeuTyrValLeuLysAsnThrGlu.....ProAlaLeuLeuGln 1270
13050 ACGTGACCTGGGCTTCGACGACGACTTCTCGCTGCACGCAATGGGCGCG 13099
1271 ArgTyrAlaThrAspLeuThrLeu 1278
13100 GCATGATGTCTGCTCGACTGTC 13123

seq_name: /cgn2_6/plodata/2/ina/5A_COMP.seq:US-08-764-233A-1
seq_documentation_block:
: Sequence 1, Application US/08764233A
: Patent No. 5716849
: GENERAL INFORMATION:
: APPLICANT: Ligon, James M.
: APPLICANT: Schupp, Thomas
: APPLICANT: Beck, James J.
: APPLICANT: Hill, Dwight S.
: APPLICANT: Neff, Snezana
: APPLICANT: Ryals, John A.
: TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Clba-Gelby Corporation
: STREET: 520 White Plains Road, P.O. Box 2005
: CITY: Tarrytown
: STATE: NY
: COUNTRY: USA
: ZIP: 10591
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/764,233A
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/729,214
: FILING DATE: 09-OCT-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/258,261
: FILING DATE: 08-JUN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Meigs, J. Timothy
: REGISTRATION NUMBER: 38,241
: REFERENCE/DOCKET NUMBER: 1506/CIP6
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (919) 541-8587
: TELEFAX: (919) 541-8689
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 49377 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Sorangium cellulosum
: IMMEDIATE SOURCE:
: CLONE: p98/1, p1J3, and pVKM15
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: 383..760
: OTHER INFORMATION: /product="SorR"
: OTHER INFORMATION: /note="This gene encodes a protein that is highly homolog

```



```

295 snPheTyRPhEaSPLeuAaSPSeRSeTMeTylSGlLeuLeuArgAla 311
      :::::::::::::::::::::::::::: |||
26642 .....GAGCTCGCGCTCCGACTCCAGACTACCTCGCTCC 26680
      |||
312 HisGlyThrHisProAlaIleSerThrLeuAla.....Argse 324
      ||| ::::::::::||| ||| |||
26681 TGGGGCGCAGAG.....CTCTCCACCGCGCGGTCAACAGCCCGCAGGGC 26724
      ::::::::::||| ::::::::::|||
324 rAlaIlePheSeRValThrTyRProSeRProAspIlePheLeuValIleL 341
      ::::::::::||| ::::::::::|||
26725 TACCTCGTATCCGCGAGCCCGCGCTCGACGCGCTCTC..... 26767
      |||
341 ySlEuglUySValLeuGlnGlnGlyAspIleSeRGlYcYsGluPro 357
      |||::: ::::::::::
26768 .....GACGTCTCCACCGCGCCACCAAGGTG 26791
      ::::::::::
358 TyRMeTValLeuUySGluValAspThrAlaIleAsnUySGluUySGlu 374
      ::::::::::||| |||::: ::::::::::
26792 TTGCGCCGCAAGATCCGCTCGACTACGCTCCCACTCCGCGCCAGATGA 26841
      ::::::::::|||
374 uLySLeuArg.....LeuAlaIaGlnPheCysThrArgLeuGlyA 389
      ::::::::::|||
26842 CGCGGTCCAGAGAGCTCGCGCA..... 26866
      |||
389 rGlyThrMetProPheAlaTrpThrAlaValHisLeuAlaAsnIleVal 405
      |||
406 SeRSeRAlaGlyGlnLeuAaSPaRgAsPSeRAsPSeRglUySGluYrAr 422
      ::::::::::||| ||| |||
26885 CCTCGGACGCGCGAGCTCCTTTATTGACCGCTCACCGCGCACAGGCT 26934
      ::::::::::|||
422 gProAlaTrpThrAspArgArgArgGlyProGlnAspArgAlaSeR 439
      ::::::::::|||
26935 CGAC.....GGCTCGC 26945
      ::::::::::
439 eRgLyAspAspAlaCysSeRSeRglYpHeArProAlaThrLeuThr 455
      ::::::::::||| ::::::::::|||
26946 AGCTCGAGCGCGCTGCTGATCGAAACCTCGCGCAACCGCTCTG... 26992
      |||
456 ValThrAsnPhElySGlnGluAlaGluArgLeuSeRAspGluSple 472
      |||::: ::::::::::||| ||| |||
26993 .....TTCTGAGCGCGACCGAGCGGCTCTCGAC...GATGG 27027
      ::::::::::
472 uPheUySPhELeuAlaSPMeTArg...ArgProSeRSeRLeuLeuArg. 487
      ::::::::::||| |||
27028 GCATCGGCTTCTCGGTGAGGTCAAGCCCCCATCCCGTCCACGCTGCGCC 27077
      ::::::::::|||
488 ArgLeuArgProValThrAlaGlnLeuUySLeAspIleSeRProAlaP 504
      |||
27078 TCCGCGAGACTCGCGAGCGCTCACCGCTCGATCCGCTGCTGCGGCTCC 27127
      |||
504 oGluAsnProHisPheCysLeuSeRPro.....GluLeuUHisIle 518
      |||
27128 ATTGAGAGAGAGAGCACTGCGCGCTGCTCTCTCTGCGCGGA 27177
      ::::::::::
519 LySProTyRProAspProArgGlyArg..... 527
      |||
27178 GCTGCTTACCGAGGCGCTCGCGCTCGACTCGAAGACTTCTTGCGCGCT 27227
      |||
528 .....ProThrLySGluIleLeuGluPheProAla 538
      |||
27228 ACGGTCCCGCAAGGTCTCCCTCCAC.....CTACCCCTTCCAGCGA 27271
      ::::::::::
538 rGglUValTyRAlaProHisThrSeRtyRArgAsnLeu..... 550
      |||
27272 GAGCGGTTCGCTCGAGCTCTCCAGGAGAGGCTTCCGAGCTGCGCT 27321
      ::::::::::
551 LeuTyRValTyRProHisSeRLeuAsnPhSeRSeRArgGln...GlySer 566
      ::::::::::||| ::::::::::|||
27322 CCGGAGGCGCTGAGCTCGGCGGACCAATCCGCTGCTGCGCGCGCTCG 27371

```

```

567 ValArgAsnLeuAlaValArgValGlnTyRMeTThrGlyUySAsPProSe 583
      |||
27372 CCTTGGCCGACCGCGGTGGCTTCTCT...TTACAGGCGGCTCTCCCTC 27418
      |||
583 rGln.....AlaLeuPro..... 587
      |||
27419 GCAGAGACACCGCTGGCTCGAAGGCCCATCGCTTTCGGCACACCATCT 27468
      |||
588 .....ValIlePheGlyLySSeRSeRSeRglUySGluYrArgGlu 601
      |||
27469 ACCGGGCAACCGGCTTCTTCGAGCTCGCCCTCGACGTCCGCCACCGCGTGC 27518
      |||
602 AlaPheThrProValValTyRHisAsnUySSeRProGluPheTyRGlU 618
      |||
27519 GCCTCGACACCGTC.....GAAGA 27537
      |||
618 uPheUySLeuHisLeuProAlaCysValThrGluAsnHisLeuLeuP 635
      |||
27538 GCTACAGCTCGAGGCGCCCTCTGCTCTCCATCGAGACACCGCTCTC. 27586
      |||
635 heThrPheTyRHisValSeRcYsGlnProArgProGlyThrAlaLeuGlu 651
      |||
27587 .....CTCAG 27592
      |||
652 ThrProValGlyPheThrTrpIleProLeuLeuGlnHisGlyArg.... 666
      |||
27593 ATCTCGCTGGG.....CCGTTGACGACGACGACGACGAGGCG 27630
      |||
667 .....LeuArgThrGlyProPhe. 672
      |||
27631 GCTCTCTTCCATAGCCGACAGAGAGAGCGGCTTCAGATGGCCCTTGA 27680
      |||
673 .....CysLeuProVal 676
      |||
27681 CTCGCGACGACGAGGCTCTCTCGCGCGGACCCCATCCCTCTCGCC 27730
      |||
677 SeRValAspGlnProProProSeRtyRSeRValLeuThrProAspValAl 693
      |||
27731 GATCTCCAGAGTGGCTCTCCCTCGAGTGCATC.....CCGCTGA 27771
      |||
693 aLeuProGlyMeTArgTyRValAspGlyHisLySGlyVal...PheSeR 709
      |||
27772 CCTCGAGGCTCTTACGCAACCTCGCCACCTCGGGCTTGGCTTACGGCC 27821
      |||
27822 CCGAGTTCAGGCGCTCGCTCGCTTACAGCGCGGAGAGAGCTCTTT 27871
      |||
726 LySPhePheThrLeuValHisValLeuGlnGluGlyAlaPheProPheAr 742
      |||
27872 GCCGAAGCCMACTCCCGAMACGGCGCGAAGATGCGCGCGGTTTGC 27921
      |||
742 gLeuUyS.....AspThrValLeuSeRglUySValGlu 755
      |||
27922 CTTCCACCTCGCGCTGCTGACAGCGCCCTGCATGCATGGCTTTGAG 27971
      |||
755 InGluLeuArgAlaSeRLeuAla.....Ala 763
      |||
27972 ACGAGAGAGAGGAGAGGTCGCTGCGCTTCTGAGGAGGAGTCTCG 28021
      |||
764 LeuArgLeuAlaSeRProGluProLeuValAlaPheSeRHisValLe 780
      |||
28022 CTGCGCTCGGTGGTGGCCACCACTTGGCGCGCTTCCAC..... 28063
      |||
780 uAspLySLeuValArgLeuValIleArgProProIleSeRglUyGlnI 797
      |||
28064 .....CTGCCAAGGTTAAATCTCCGCT 28088
      |||
797 leValAsnLeuGlyArgGlyAlaPheGluAlaMeTAlaHisValSeR 813
      ::::::::::
28089 CGATGCTCTGCGCGAGCGCGAGGAGTACCTCTTGGCTCGGTGCAAGCG 28138
      ::::::::::
814 LeuValHisArgSeRLeuGlnAlaIaGln.....AspAlaArgG 827

```





```

996 .....TTTGAGTGGCTGTGATGATGT 1018
456 lThrAspPhePheLysGlnGlnAlaGluArgLysSerAspGluAspLeu 473
|||||.....:|||||
1019 AACAGATATAATAAT.....GGAAGATGATGATGAAGATTAAC 1059
473 heLysPheLeuAlaAspMetArgArgProSerSerLeuLeuArgArgLeu 489
|||||.....:|||||
1060 ACCATTTCATTCC.....TTT 1076
490 ArgProValThrAlaGlnLeuLysIle.....AspIle 500
:|||||.....:
1077 CACCGGTGGAGGAGGAATGACTCTTCACAGCTGTTATAACAAGCT 1126
500 eSerProAlaProGluAsnProHis.....P 509
: ||| |||
1127 CATCGGTGCCAAGAAGTCAACACCAAGGGCAGGCTTGTGGTAACT 1176
509 heCysLeuSerProGluLeuLeuHis.....IleLysProTyrProAsp 523
: ||| ||| :|||
1177 TGAATTACTTCTCGAGATATCCATCAGATCCGAAAGAGTTTCCGCAT 1226
524 ProArgGlyArgProThrLysGlnIleLeuGluPheProAlaArgGluVal 540
: ||| ||| :|||
1227 TTAGTGCACGAGCACACAGCTGTGCTCGAAAAACAGGCTTCCGAGAT 1276
540 lTyrAlaProHisThrSerTyrArgAsnLeuLeuTyrValTyrProHis 557
: ||| :||| :|||
1277 AATCATGCGCT...GGTATGTTGGAATGATATCTATGTAACTATGTTGC 1323
557 eLysLeuAsnPro...SerSerArgGlnGlySerValArgAsnLeuAlaVal 572
: :||| :||| :|||
1324 AAGGAGATTTTGTATTAAGACGCAAAACAAACGCAAGCAACGTTGAGTGC 1373
573 ArgValGlnTyrMetThrGlnGluAspProSerGlnAlaLeuProValIle 589
|||||.....:|||||
1374 ACAGGTCT...GTGACAGATGAGATGGGAACGATTGACAGCTGTGAT 1420
589 ePhe.....GlyLysSerSerCysSerGluPheThrArgGluAlaIle 603
||||| ||| :||| :|||
1421 TTTCGGGGTGGTGGTGAAGCGATTTCAAGTGCANA..... 1460
603 heThrProValIleValTyrHisAsnLysSerProGluPheTyrGlnGluPhe 619
: :||| :||| :|||
1461 ..TCTGTGATTACTACCAAGTAAACACACGCGTGTGAGACTGT 1508
620 LysLeuHisLeuProAlaCysValThrGluAsnHisLeuLeuPheThr 636
||||| :||| :|||
1509 AAGTGGCCATTCCTCCATCGAGAGCGTTAAACCGCAGTCCGCTTTC 1558
636 rPheTyrHisValSerCysGlnProArgProGlyThrAlaLeuGluThrP 653
||||| ||| |||
1559 CTTCGCCACAGGTCAATCAGAGACTCTAAGATTAATCT...GAGAAA 1605
653 roValGlyPheThrTrpIleProLeuLeuGlnHisGly.....ArgLeu 667
: :||| :||| :|||
1606 TATTGCACTAGCATTTGTCAAGCTGATGATAGATGATGATCAACCTG 1655
668 ArgThrGlyProPheCysLeuProValSerValAspGlnProProPhe 684
||| ||| |||
1656 CAGAGCGGAGAGACAGCATCTTATGCTCTATAAGCCGGAAGCGAAGAGCT 1705
684 rTyrSerValLeuThrProAspValAlaLeuProGlyMetArg..... 698
: :||| :||| :|||
1706 GGAAGATGCTGCACG...TACTTGAGTCTGCCCTCCAGAAAGGAGAGT 1752
699 ..TrpValAspGlnHisLysGlyValPheSerValGlnLeuThrAlaVal 714
: ||| :||| :|||
1753 TCGAAGAAAAAGGGCCATCGCCACCGCAAGACGATGACAGCTTGGG 1802
715 SerSerValHisProGlnAspProTyrLeuAspLysPhePheThrLeuVal 731
||| :||| :|||
1803 ACCTGCACACATTACGACAGACTCTTC.....CAGATCTCCACGCTGCT 1846

```

```

731 l.....HisValLeuGlnGluGlyAlaPhePro 741
|.....:|||||
1847 GTGCTCCACCAACTGACTGAAACGTGGACCTTGTGGGCTTGTGAAT 1886
741 heArgLeuLysAspThrValLeuSerGlnGluAsnValGlnGlnGluLeu 757
: ||| :||| :|||
1897 GCGCTCCACACACGCGCTGCAGAG..... 1925
758 ArgAlaSerLeuAlaAlaLeuArgLeuAlaSerProGluProLeuValAl 774
: ||| :||| :|||
1926 .....AACTGAGCGCAGCTGATGAAGTGCATGCTGTGAAGTGA 1969
774 aPheSerHisValLeuAspLysLeuValArgLeuValIleArgPro 791
||| :||| :|||
1970 GTTCTTTCAGACACAGCTTGATGCCCTCTTCACATCATGATGAG... 2015
791 rolIleLeSerGlyGlnIleValAsnLeuGluArgGlyAlaPheGluAla 807
: :||| :||| :|||
2016 .....AACTCAGAGAGTGAGACTTTGACAGCTTAGCTTGTGATGCT 2057
808 MetAlaHisValIleSerLeuValHisArgSerLeuGlnAlaAlaGlnAs 824
: :||| :||| :|||
2058 CTGGATTATTACCATGAGCTGATT.....GCTGATACAAA 2092
824 PALArgGlyHisCysProGlnLeuAlaAlaTyrValHisTyrAlaPheA 841
: ||| :||| :|||
2093 ATTTCAGCATTTTAATCCTGTTTGGAACTTACATTACAAACACTTT 2141
841 rGluProGlyThrGluProSerLeuProAspGlyAlaProProValIle 857
2141 ..... 2141
858 ValGlnAlaAlaThrLeuAlaArgGlySerGlyArgProAlaSerLeuTyr 874
: :||| :||| :|||
2142 .....AGTGCACGCTTAGCC.....TA 2158
874 rLeuAlaArgSerLysSerIleSerSerSerAsnProAspLeuAlaValA 891
|.....:|||||
2159 CACGAGTGTGACAAA...GTGTTGAAGAACTACGTGACGTCGTCTGAGA 2205
891 lAProGlySerValAspAspGluVal..... 899
||||| :||| :|||
2206 AGCCGGGA...GTAAATGAGCAGCTGTACAAAGCCATGAAGCGCTGAA 2252
900 .....SerArgIleLeuAlaSerLysLeu 908
||||| :||| :|||
2253 TCCATCTTCAAGTTCAATCGTGGCTCCAGATCTCTGTCAATCACTGTA 2302
908 uHisGlnGluLeuAlaLeuGlnTrpValIleSerSerSerAlaValArg 925
: :||| :||| :|||
2303 TGAAGAACAGGAGAGCGCTGACTCGTG.....G 2331
925 lValAlaIleLeuGlnHisAlaTrpPhePheGlnLeu.....MetVal 939
: ||| :||| :|||
2332 AATTTTGTCTGAG.....CTCTCAGGTCCATCAATGACATGATG 2372
940 LysSerMetAlaLeuHisLeuLeuGlyGlnArgLeuAspThrProAr 956
: :||| :||| :|||
2373 AGCAGCATGTCA..... 2384
956 gLysLeuArgPheProGlyArgPheLeuAspAspIleThrAlaLeuValG 973
||| ||| :|||
2385 .....GACACAGACCGTCGGGTCA 2403
973 LysSerValGlyLeuGluValIleThrArgValHisLysAspValGluLeu 989
: :||| :||| :|||
2404 AGGGGCGACAGCTGAATACTTACCAACGATCGTCAACGATGGAATTG 2453
990 AlaGlnHisLeuAsnAlaSerLeuAlaPhePheLeuSerAspLeuLeu 1006
2453 ..... 2453

```

1006 rLeuValAspArgGlyPheSerLeuValArgAlaHisTyrLysG 1023  
1023 InValAlaThrArgLeuGlnSerSerProAsnProAlaIleLeuThr 1039  
2460 .....GATCCAAAGAGCTCAGCAAA 2480  
1040 LeuArgMetGluPheThrArgIleLeuCysSerHisGluHisTyrValTh 1056  
2481 ATGTTTACTGAATTTC.....AT 2497  
1056 rLeuAsnLeuProCysCysProLeuSerProProAlaSerProSerPro 1073  
2498 CCTCAATGTTCC..... 2510  
1073 eValSerSerThrThrSerGlnSerSerThrPheSerSerGlnAlaPro 1089  
2510 ..... 2510  
1090 AspProLysValThrSerMetPheGluLeuSerGlyProPheArgGlnG 1106  
2510 ..... 2510  
1106 nHisPheLeuAlaGlyLeuLeuLeuThrGluLeuAlaLeuAlaLeuGln 1123  
2511 .....ATGGGCTTGGTACCATCCAGAACTCTACTGCTTG.... 2546  
1123 roGluAlaGlnGlyAlaPheLeuLeuHisLysLysAlaIleSerAlaVal 1139  
2547 .....ATCGAAATCGTC 2558  
1140 HisSer...LeuLeuCysGlyHisAspThrAspProArgTyrAlaGluAl 1155  
2559 CACAGTACCTCTTCACACAGCATGAC..... 2585  
1155 aThrValLysAlaArgValAlaGluLeuTyrLeuProLeuLeuSerIleA 1172  
2586 .....TGCAGAGAGATCCCTGCTTCCCATGATGAC.... 2614  
1172 IaArgAspThrLeuProArgLeuHisAspPheAlaGlnGlyProGlyGln 1188  
2614 ..... 2614  
1189 ArgSerArgLeuAlaSerMetLeuAspSerAspThrGlnGlyGlnGlyAs 1205  
2615 .....CGATCAGCTCAAGTACCATCTGGAGAGACAGAGACCTGGAGG 2658  
1205 rIleAlaGlyThrIleAsnProSerValAlaMetAlaIleAlaGlyGlyP 1222  
2659 CCTGCTGTCAAGCTCTCAGCCACAT.....CCTGGAGGTG 2693  
1222 roLeuAlaProGlySerArgAlaSerIleSerGlnGlyPro..... 1235  
2694 CTGTACAGAAAGACGTGGGGCCAAAGAGGAGAGTCCAGATTATCAT 2743  
1236 .....Pro.ThrAlaSerArgAlaGlyCysAlaLeuSerAlaG 1248  
2744 GGAGAAACTTCTCGGACGCTGAACCGAACCGTATTTCATGGAGAGAG 2793  
1248 LuSerSer.....ArgThrLeuLeuAlaCysValLeuTyrValLeuLys 1262  
2794 ATTGTGAACATCATTTGGAAACTCTGCTGCTTCATGACAGCTATTATTACGA 2843  
1263 AsnThrGluProAlaLeuLeuGlnArgTyrAlaThrAspLeuThrLeuPr 1279  
2844 CAATTCGAA..... 2852  
1279 oGlnLeuGlyArgLeuLeuAspLeuLeuTyrLeuCysLeuAlaAlaPheG 1296  
2853 .....GATTACCATTTATGCCCACTTGATCAAG.... 2879  
1296 LuTyrLysGlyLysAlaPheGlnArgIleAsnSerLeuThrPheLys 1312

2880 .....ACTTTGGG 2888  
1313 LysSerLeuAspMetLysAlaArgLeuGlnGluAlaIleLeuGlyThrI 1329  
2889 AAA.....ATGAGACATGATGTGGTAGATTTCCTAATGAAGAACTT 2929  
1329 eGlyAlaArgGlnGluMetValArgArgSerArgGluArgSerProPhe 1345  
2930 CATCATGTTTAAAGAACCTCAT.....GGAAGAACCTTTAACCCCTTCG 2973  
1346 .....GlyAsnProGluAsnValArgTyrAlaLysSerVal 1357  
2974 ACTGGGTGATCATGAACATGTCGCAAAATTAAGTTCTTCGAGACAAAT 3023  
1358 ThrHisTyrLysGlnThrSerAspArgValAspLysThrLysAspGlu 1374  
3024 AATCAGTATGACAGATATGCTGAACAA.....AATTTCTGATCAACG 3067  
1374 tGluHisGluAlaLeuValGlnGlyAsnLeuAlaThrGluAlaSerLeu 1391  
3068 CAACCTTGAGCTACAGCTGTGGAAACACTTTCACCTGCT.....G 3111  
1391 aValLeuAspThrLeuGlnIleIleValGlnThrValMetLeuSerGlu 1407  
3112 TTGCTTTCTTACTCAAGAGTCCCTGCAACTGGAGAAATTTTCAAGTCC 3161  
1408 AlaArgGluSerValLeu.....G 1414  
3162 AAGAGAGCCAAATCCTTAAACAATGAGATATGAGAGACAGATGG 3211  
1414 yAlaValLeuLysValValLeuTyrSerLeuGlySerAlaGlnSerAla 1431  
3212 CTTTGAATCAGACAGACATGTGGTCAACCTTGT..... 3245  
1431 euPheLeuGlnHisGlyLeuAlaThrGlnArgAlaLeuValSerLysPhe 1447  
3246 .....CAACCAAGATTAAGTTTCATCCGAAATGTTGGCCCATTA 3287  
1448 ProGluLeuLeuPheGlnGluAspThrGluLeuCysAlaAspLeuCysLe 1464  
3288 TTAGAATGACATTAAATCCGAGAGCGAGCTG..... 3320  
1464 uArgLeuLeuArgHisCysGlySerArgIleSerThrIleArgThrHisA 1481  
3321 .....CGCAAGCCACCATTC..... 3335  
1481 laSerAlaSerLeuTyrLeuLeuMetArgGlnAsnPheGluIleGlnHis 1497  
3336 .....CCCATCTTCTTTGATATGATGACAGTGTGAATTCATTCGACCCGA 3380  
1498 AsnPheAlaArgValLysMetGlnValThrMetSerLeuSerSerLeuVal 1514  
3381 ACCTTCCAAAATGTTTGAATAATGATCATCAACCAAGCTGATCAGAAGT 3430  
1514 L.....GlyThrThrGln.....AsnPhe 1521  
3431 CGAAGGAGGACAGAGAGACAAACAGTATTAATTTGATAAAATCC 3480  
1521 eGlnGluHisLeuArgArgSerLeuLysThrIleLeuThrTyrAlaGlu 1537  
3481 TTCTGGAACACTGCAGAGAG.....CAC 3503  
1538 GluAspMetGlyLeuArgAspSerThrPheAlaGlnGlnValGlnAspLe 1554  
3504 AAATACCTCGCCAAAGACAGAGAAACTTTTGAAACGCTGTGCGCTT 3553  
1554 uMetPheAsnLeuHisMetIleLeuThrAspThrValLysMetLysGluH 1571  
3554 AATGAAAGCTTTTGATTTATGAACCATCATG.....C 3588  
1571 iSGlnGluAspProGluMetLeuIleAspLeuMetTyrArgIleAlaArg 1587

```

3589 ACGACGAGACAAAGAAAACCCATGAGCTGCACCTCAATGCTCGAAT 3638
1588 GlycerProaspLeuArgLeuThrThrLeuGlnAs 1602
3639 TTCTACAAAGAAATTTGAAAGAGAAATGATATTAAGGTATTTGTCAA 3688
1602 nmetAlaGlyLysHisAlaGluGlyAsnHisAlaGluAlaAlaGlnC 1619
3689 GCTCTGTGACCTGCACAAGAGGTGTGATTAACACCGAAGCGCTTACA 3738
1619 ysmetValHisAlaAlaAlaLeuValAlaGluThrLeuAlaLeuGlu 1635
3739 CTTGCTT.....CTCCATGCAAAAGCTTTTAAGTGGTCGAG 3776
1636 AspGlnArgHisLeuProValGlyCysValSer..... 1646
3777 GATGTG.....TGTGTGCCCCACTCACCCAGCGCGGA 3808
1647 ....PheGlnAsnIleSerSerAsnValLeuGluGluSerAlaIleSerA 1662
3809 CGGGTACACGACCCACGACGAGGAGCTGAAGAGCAGCTGTACAGG 3858
1662 spARPIleLeuSerProAspGluGluGlyPheCysSerGlyLysHisPhe 1678
3859 AAATCATC.....CACTACTTCGACAAAGGCAAGATGTGC 3893
1679 ThrGluLeuGlyLeuValGly.....LeuLeuGluGlnAlaAlaGly 1693
3894 GAGAGGCGCATTCCTTGGCGAAGAGACTAGCCGACAG..... 3932
1693 rPheThrMetGlyGlyLeuThrGluAlaValAsnGluValThrLysAsn 1710
3933 .....TATGAG.....AACGAATGTT..... 3950
1710 euIleProIleLeuGluAlaHisArgAspThrLysLysLeuAlaAlaVal 1726
3951 .....GATTATGAGCAACTCAGCGAATTTG 3974
1727 HisGlyLysLeuGlnGluAlaPheThrLysIleMetHisGlnSerSerG 1743
3975 CTGAAGAAACAGCGCTGATTATGAAACATGCTCAAGTGAATC.... 4019
1743 yTrpGluArgValPheGlyThrThrPheArgValGlyPheThrGlyAla 1760
4020 .....AGGCCAAGCCTGACTATTTGCTGGCTACTAGCGACAG 4062
1760 AspPheGlyAsp...LeuAspGluGlnGluPheValThrLysGluProSer 1775
4063 GGTTCGCCACATTCCTCGGGGAAAATTTTCATTACCGAGGAAAGAG 4112
1776 IleThrLysLeuAlaGluIleSerHisArgLeuGlnGluPheThrGln 1792
4113 TATGAGCCCGCGAAGATTTTGAAGCTGCGCTTAACCTAGTTCCAA 4162
1792 uArgPheGlyAspAspValValGluIleIleLysAspSerThrProVal 1809
4163 C.....GCCGAGAAATGAAACAACTCTCCACACAG 4194
1809 sPlySerLysLeuAspSerGlnLysAlaThrIleGlnIleThrVal 1825
4195 GCGACGATATTAAACTCTCTGCGCAGATATTTCAGTCTTCACAGTG 4244
1826 GluProThrPheAspThr.....TyrlGluLeuLysAspAr 1837
4245 AAGCCCAAACTGCATGCGCTCAAGTTTCACAGCGCCATGTACAGCA 4294
1837 gValThrThrPheAspArgAsnThrGlyLeuArgThrPheLeuPheCys 1854
4295 GATTGTAAAGTTTTCAGGCTGAACGAGGTCAGCGATTGGAATATCTC 4344
1854 hrProPhe.....ThrProAspGlyArgAlaHisGlyGlu 1865
4345 GGGCAATCCGAGAGGAGAGAAACCCAGAC.....ATGAA 4382

```

```

1866 LeuProGluGlnHisLysArgLysThrLeuLeuSerThrAspHisAlaPh 1882
4383 TTTCGCAATATGTCATCGAGAAACCATATATACAACTGCATATTAATT 4432
1882 eProThrIleLysThrArgIleArgValCysHisArgGluGluThrVal 1899
4433 ACCTGCAATTTTAAGGTGTGTGAGTCAAGTCTGTTTCATGATGGAAA 4482
1899 euThrProValGluValAlaIleGluAspMetGlnLysLysThrArgGlu 1915
4483 TCAGCCCTGAGATGCCATCGAGACCATGACCTGACGACAGACAG 4532
1916 LeuAlaPheAlaThrGluLysProProAspAlaLysMet..... 1929
4533 ATCAACAGCATGTGTGACGACGACCTGATGACCCAGCTCCCATCAA 4582
1930 ....LeuGlnMetValLeuGlnGlySerValGlyProThrValAsnGln 1945
4583 CCCGCTCTCATGCTCCCTGACGCGCATGCTGAGCCAGCTGTATGGGG 4632
1945 lProLeuGluValAlaGlnValPheLeuAlaGlu.....IleProGlu 1959
4633 GCTTCGCAAACTACGAAAGGCTTCCTTACAGACCGGTACCTGCAGAG 4682
1960 AspProLysLeuPheArgHisHisAsnLysLeuArgLysCysPheLysAs 1976
4683 CAOCCTGAGCCCGCATGAAAGATCGAAGAGCTC.....AAGA 4720
1976 pPheCys..... 1978
4721 CCGATTGCTGTGCGAGATTCCTTTCTGCGCAGAGCATCAATCATG 4770
1979 ..LysLysCysGluAspAlaLeuArgLysAsnLysAlaLeuIleGlyPro 1994
4771 GAGACAAAGTCACGAGGAGCATGAG..... 4796
1995 AspGlnLysGluThrHisArgGluLeuGluArgAsnThrCysArgLeuAr 2011
4797 .....CCGTCACAGAGAGATGAGCGCTTTCAAACACTGAA 4837
2011 gGluAlaLeuGln 2015
4838 GGAAGAGTGGAG 4850

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-971-988-1
seq_documentation_block:
; Sequence 1, Application US/08971988
; Patent No. 5786461
; GENERAL INFORMATION:
; APPLICANT: Michiyuki MATSUDA et al.
; TITLE OF INVENTION: cDNA OF DOCK180 GENE AND DOCK180 PROTEIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/971,988
; FILING DATE: 17-NOV-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/588,985
; FILING DATE:

```



ATTORNEY/AGENT INFORMATION:  
NAME: Warren M. Cheek, Jr.  
REGISTRATION NUMBER: 33,367  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-8850  
TELEFAX:  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6519 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL:  
ANTI-SENSE:  
FRAGMENT TYPE:  
ORIGINAL SOURCE: spleen cell of homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 24..5619  
US-08-971-988-1

alignment\_scores:  
quality: 220.00 length: 2105  
ratio: 0.260 gaps: 93  
Percent Similarity: 40.238 Percent Identity: 18.432

alignment\_block:

US-09-736-968A-2 x US-08-971-988-1 ..

Align seg 1/1 to: US-08-971-988-1 from: 1 to: 6519

```
78 pheprolaaspaapleugluLeuLeuLeuInProArGluCysArgTh 94
|||||
201 TTTCCTGCTCATATTCATCTT.....AAGAGCGGATAGT 238
94 rThrGluProGly.....IleProLysAspGluLysLeuA 106
|||
239 TGAAGCAAAAGGCAACATGAACAAGTATCCGGGTGACCTCCCGCTCA 288
|||
106 spAlaGluValArgAlaVal.....Glu 114
|||
289 TCCAGGAGTACACGACACTCCGAGAGTGTCACCATCTGGAGGAG 338
|||
115 MetYrIleGluAsp..... 119
|||
339 CTCATCGTGCAGATACAGGAGATGTTGCAAGTGTGGCGCACATGAT 388
|||
120 .TrpValIleValHisArgArgTyrgInTyrlLeuSerAlaAlaTySerP 136
|||
389 CTATGACCTTATTAATGCGATCACAATCTTCTGGAAGT..... 431
|||
136 rovalrThrAspThrgInArgGluArgGluLysGlyLeuPro..... 150
|||
432 ..CTGCCTCAGAGTGAAGTCAAGAGTGAAGAGAGTCAACAGCCAA 479
|||
151 .....ArgGluValPheGluGluInAspAlaSerGlyAspGluAr 163
|||
480 ATTGATTATGCAACAGATTTCTAGATTGGACCTGCTGGTTAGAT.. 527
|||
163 gserGlyProGluAspSerAsnAspSerArgArgGlySerGlySerProg 180
|||
528 .....GAAGATGGGAATATTG..... 545
|||
180 lAspThrProArgSerSerGlyAlaSerSerIlePheAspLeuArgAsn 196
|||
546 .....GATCCAGATTAATACACAGATTAAGTCTTCAGAGCTCATGAA 590
|||
197 leuAlaAlaAspSerLeuLeuProSerLeuGluArgAlaAlaProGl 213
|||
```

```
591 ATAGCTTCT.....AACAACTGGAGAAAGCTTACAAGAGA 628
213 uAspValAspArgArgAsnGluThrLeuArgArgGluHisArgProProA 230
|||
629 AAAATCTCAAAAGCAGAAACATAGATATTAAACAGAACCAAGTTGCTG 678
230 lAleu.....LeuThrLeuTyrrProAlaProAspGluAspGluAlaVal 244
|||
679 CAACCCCTTCTCTGGCCCTGTT..... 701
245 GluArgCysSerArgProGluProProArGluHisPheGlyGluArgI 261
|||
702 .....GT 703
261 eleuValCysLeuSerLeuLysPheGluIleGluIleGluProIleP 278
|||
704 GAACCTCAAAAGTGTGTTGTAATAATAGAGAAATGCTGAAGTC.... 749
278 heGlyIleLeuAlaLeuTyrrAspValArgGluLysLysLysIleSerGlu 294
|||
750 .....CTCATGTCTATATGACCCCTGTGGAGTCCAAATTCATCATGAG 794
295 AsnPheTyrrPheAspLeuAsnSerAspSerMet...LysGlyLeuLeuAr 310
|||
795 AACTACCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 844
310 gAlaHisGlyThrHisProAlaIleSerThrLeuAlaArgSerAlaIleP 327
|||
845 ATACATTAATTTGCGCGCGCTGTTACTGACTCGGA..... 881
327 heSerValThrTyrrProSerProAspIlePheLeuValIleLysLeuGlu 343
|||
882 .....AGC 884
344 LysValLeuGluGlnGlnLysPheLysSerGluCysGluProTyrrMetVa 360
|||
885 AAAGACTGTGAAGGAGCAAAATCATGTTTGTCTGTCAG..... 923
360 lLeuLysGluValAspThrAlaLysAsnLysGluLysLeuGluLysLeuA 377
|||
923 ..... 923
377 rGluAlaAlaGluGlnPheCysThrArgLeuGlyArgTyrrArgMetPro 393
|||
924 .....ATTGTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 950
394 PheAlaThrThrAlaValHisLeuAlaAsnIleValSerSerAlaGly 410
|||
950 ..... 950
410 nLeuAspArgAspSerAsp.....SerGluGlyGluArgArgP 423
|||
951 .....AGGAGACAAACACACAGAAACTGACCTGGGCTGCTGCTGCTG 993
423 roAlaThrThrAspArgArgArgGlyProGluAspArgAlaSerSer 439
|||
994 CT..... 995
440 GlyAspAspAlaCysSerPheSerGlyPheArgProAlaThrLeuThrVa 456
|||
996 .....TTTGAGTGTGCTGTGATGATGT 1018
456 lThrAsnPhePheLysGlnGluAlaGluArgLeuSerAspGluAspLeuP 473
|||
1019 AACAGATTAATAAAT.....GGAAGAGTGAAGATGAAGATGAAGC 1059
473 heLysPheLeuAlaAspMetArgArgProSerSerLeuLeuArgArgLeu 489
|||
1060 AGCATTTTCATTCG.....TTT 1076
490 ArgProValThrAlaGluLeuLysIle.....AspIle 500
|||
1077 CAGCCGCTGCGAGGAGAAATGACTCTTCAGACTGTATTAACAACAAGT 1126
|||
```



```

|||||:|||||
2498 CCGAATGTTCC..... 2510
1073 eValSerSerThrThrSerGlnSerSerThrPheSerSerGlnAlaPro 1089
2510 ..... 2510
1090 AspProLysValThrSerMetPheGlnLeuSerGlyProPheArgGlnGln 1106
2510 ..... 2510
1106 nhISpHeuAlaGlyLeuLeuLeuThrGlnLeuAlaLeuAlaLeuGlnP 1123
|||||:|||||
2511 .....ATGGGCTGCTGTCACATCCAGAACTCTACTGCTTG.... 2546
1123 roGlnAlaGlnGlnAlaPheLeuLeuHisLysAlaAlaLeuSerAlaVal 1139
|||||:|||||
2547 .....ATCGAAATCGTC 2558
1140 hISer...LeuLeuCysGlyHisAspThrAspProArgTyrAlaGlnAl 1155
|||||:|||||
2559 CACAGTGACCTCTTCACACAGCATGAC..... 2585
1155 arThrValLysAlaArgValAlaGlnLeuTyrLeuProLeuLeuSerIleA 1172
|||||:|||||
2586 .....TGCAGAGAGATCCTGCTTCCCATGATGAC..... 2614
1172 laArgAspThrLeuProArgLeuHisAspPheAlaGlnGlyProGlnGln 1188
2614 ..... 2614
1189 ArgSerArgLeuAlaSerMetLeuAspSerAspThrGlnGlyGlnLys 1205
|||||:|||||
2615 .....CGATCAGCTCAAGTACCATCTGGAGACAGAGACCTGGAGG 2658
1205 pIleAlaGlyThrIleAsnProSerValAlaMetAlaIleAlaGlyLysP 1222
|||||:|||||
2659 CCTGCTGTCAGCTGCTCAGCCACAT.....CCTGGAGGTG 2693
1222 roLeuAlaProGlnSerArgAlaSerIleSerGlnGlyPro..... 1235
2694 CTGACAGAGAGAGCTGGGGCCAAACCCAGAGCAGCTCCAGATTATCAT 2743
1236 .....Pro..ThrAlaSerArgAlaGlyCysAlaLeuSerAlaG 1248
|||||:|||||
2744 GGAGAAACTTCTCCGACGCGTGAACGACGTCATTTCCATGGAGCAG 2793
1248 LuSerSer.....ArgThrLeuLeuAlaCysValLeuThrPValLeuLys 1262
|||||:|||||
2794 ATTCTGAACATCATTTGGAACCTGCTGCTGCATGACAGCTATTATTACGA 2843
1263 AsnThrGlnProAlaLeuLeuGlnArgTyrAlaThrAspLeuThrLeuPr 1279
|||||:|||||
2844 CAAATGGAA..... 2852
1279 oGlnLeuGlyArgLeuLeuAspLeuLeuTyrLeuCysLeuAlaAlaPheG 1296
|||||:|||||
2853 .....GATTACCATTTATGCCCACTGATCAAG.... 2879
1296 LuTyrLysGlyLysLysAlaPheGlnArgIleAsnSerLeuThrPheLys 1312
|||||:|||||
2880 .....ACTTTGGG 2888
1313 LysSerLeuAspMetLysAlaArgLeuGlnGlnAlaIleLeuGlnTyrIle 1329
|||||:|||||
2889 AAA.....ATGAGAGACTGATGTGTAGATTCTCTAATGGAACAATT 2929
1329 eGlyAlaArgGlnGlnLeuValAlaArgArgSerArgGlnArgSerProPhe. 1345
|||||:|||||
2930 CATCATGTTTAAGAACCTCATT.....GGAAAGAACGTTTACCCTTCCG 2973
1346 .....GlyAsnProGlnAsnValAlaGlyTyrPargLysSerVal 1357
|||||:|||||

```

```

2974 ACTGGGTGATCATGACATGGTGCAAAATTAAGTCTTCTCCGAGCAATT 3023
1358 ThrIstIrrLysGlnThrSerAspArgValAspLysThrLysAspGlnLe 1374
|||||:|||||
3024 AATCAGATATGACGATATGTGGAACAA.....AAATTCTGGATCAAGC 3067
1374 tGlnHisGlnAlaLeuValGlnGlyAsnLeuAlaThrGlnAlaSerLeuY 1391
|||||:|||||
3068 CAACCTTGAGCTACAGCTGTGGAACAACTACTTTCACTGGCT.....G 3111
1391 aValLeuAspThrLeuGlnIleLeuValGlnThrValMetLeuSerGln 1407
|||||:|||||
3112 TTGCTTCTCTTAATCTCAAGAGTCCCTGCACCTGGAGAAATTTTCAAGTCC 3161
1408 AlaArgGlnSerValLeu.....G 1414
|||||:|||||
3162 AAGAGAGCCAAATTCCTTAACAGTACGAGATATGAGAGACAGATTGG 3211
1414 yAlaValLeuLysValValLeuTyrSerLeuGlySerAlaGlnSerAla 1431
|||||:|||||
3212 CTTTGAATCAGACAGATGTGTACAACCTTGGT..... 3245
1431 euPheLeuGlnHisGlyLeuAlaThrGlnArgAlaLeuValSerLysPhe 1447
|||||:|||||
3246 .....CAACACAGATTAAGTTTCATTCAGAAATGTTGGCCCAATA 3287
1448 ProGlnLeuLeuPheGlnGlnAlaAspThrGlnLeuCysAlaAspLeuCys 1464
|||||:|||||
3288 TTGAAGATGACATTATTCGCGAGCGAGGCTG..... 3320
1464 uArgLeuLeuArgHisCysGlySerArgIleSerThrIleArgThrHisA 1481
|||||:|||||
3321 .....CCGAAAGCCACATC..... 3335
1481 laSerAlaSerLeuTyrLeuLeuMetArgGlnAsnPheGlnIleGlyHis 1497
|||||:|||||
3336 .....CCCATCTTCTTTGATATGATGACAGTGTGAATTCATTCAGCCGA 3380
1498 AsnPheAlaArgValLysMetGlnValThrMetSerLeuSerLeuVal 1514
|||||:|||||
3381 ACCTTCCAAATGTTGAAATGAGATCATCACCAAGCTGATCATGAAGT 3430
1514 L.....GlyThrThrGln.....AsnPhe 1521
|||||:|||||
3431 CCAAGAGGCGAGAGAGACAGACAGTACAAAGTGTATTGTATAAATCC 3480
1521 eGlnGlnIstLeuArgArgSerLeuLysThrIleLeuThrTyrAlaGln 1537
|||||:|||||
3481 TTCTGGAACACTGCAGGAAG.....CAC 3503
1538 GluAspMetGlyLeuArgAspSerThrPheAlaGlnGlnAlaAspLe 1554
|||||:|||||
3504 AAATACCTCGCCAAACAGAGAACTTTTGTAACCTGTTGCGCTT 3553
1554 uMetPheAsnLeuHisMetIleLeuThrAspThrValLysMetLysGln 1571
|||||:|||||
3554 AATGAAAGGCTTTGGATTATAGAACCATCATG.....C 3588
1571 IsGlnGlnAspProGlnMetLeuIleAspLeuMetTyrArgIleAlaArg 1587
|||||:|||||
3589 ACGACGAGAAACAAGAAACCGCATGAGCTGCACCGTCAATGCTGTAAT 3638
1588 GlyTyrGln.....GlySerProAspLeuArgLeuThrTyrLeuGlnAs 1602
|||||:|||||
3639 TTCTACAAAGAAATTCGAAAGAGAGAAATGTATATAGGATTTGTACAA 3688
1602 nMetAlaGlyLysHisAlaGlnLeuGlnLysnHisAlaGlnAlaAlaGln 1619
|||||:|||||
3689 GCTGTGTGACCTGCGACAGAGAGTGTATACACTACACCGAAGCGCTTACA 3738
1619 yMetValHisAlaAlaAlaValAlaGlnTyrLeuAlaLeuGlnGln 1635
|||||:|||||
3739 CATTGCTT.....CTCCATGCAAGAGCTTCTTAAGTGTGCGAG 3776
|||||:|||||

```

```

1636 AspGlnArgHisLeuProValGlyCysValSer..... 1646
1637 GATGTG.....TGCTGGCCCTCACCACCGCGGCA 3808
1647 ....PheGlnAsnIleSerSerAsnValLeuGlnIleSerAlaIleSerA 1662
1648 .....:.....:.....:.....:.....:.....:
1649 CGGCTACCGCCACCGCAGGAGCAGCTGAGGAGCAGCTTACCAAG 3858
1662 spAspIleLeuSerProAspGluGlyPheCysSerGlyLysHisPhe 1678
1663 :.....:.....:.....:.....:.....:
1664 AATTCATC.....CACTACTTGACAAGGCAAGATGTGG 3893
1679 ThrGluLeuGlyLeuValGly.....LeuLeuGlnGlnAlaIleGlyTyr 1693
1680 GAGGAGGCGCATTCCTTGGCAGAGCAGTACCGGAGCAG..... 3932
1693 rPheIleMetGlyGlyLeuTyrGlnAlaValAsnGluValTyrLysAsnL 1710
1694 .....:.....:.....:.....:.....:
1695 .....TATGAG.....AAGCAATGTTT..... 3950
1710 euIleProIleLeuGlnAlaHisArgAspTyrLysLysLeuAlaVal 1726
1696 .....:.....:.....:.....:.....:
1697 HisGlyLysLeuGlnGlnAlaPheThrLysIleMetHisGlnSerSerG 1743
1698 CTGAAAAAACAGGCTCAGTTTATGAAAAACATGTCACAGTATC..... 4019
1743 yTrpGluArgValPheGlyThrTyrPheArgValGlyPheTyrGlyAlaH 1760
1744 .....:.....:.....:.....:.....:
1745 .....AGGCCAACGCTGACTATTTCTGTGGTACTACGGGCAAG 4062
1760 IsPheGlyAsp...LeuAspGlnGlnIlePheValTyrLysGluProSer 1775
1761 :.....:.....:.....:.....:
1762 GGTTCCTCCATCTCTCGCGGAAAGTTTCATTACCGAGGAAAGG 4112
1776 IleThrLysLeuAlaGlnIleSerHisArgLeuGlnIlePheTyrThrG 1792
1777 :.....:.....:.....:.....:
1778 TATGACCCCCGGAGATTTTGAAGCTCGCTTAACCTGCTTCCAAA 4162
1792 uArgPheGlyAspAspValAlaGlnIleLysAspSerTyrProValA 1809
1793 :.....:.....:.....:.....:
1794 C.....GCGGAAATAATGACACACATGCTCCACAG 4194
1809 sPlySerLysLeuAspSerGlnLysAlaTyrLleGlnIlePheTyrVal 1825
1810 :.....:.....:.....:.....:
1811 GCGAGATATTAAAAAACTCTCGCCAGTATATCAGTCTTACAGTG 4244
1826 GluProTyrPheAspThr.....TyrGluLeuLysAspArg 1837
1827 :.....:.....:.....:.....:
1828 AAGCCCAACTGATCTGCTCTAGTTTACAGGCGCAGTCTGAGACA 4294
1837 gValThrTyrPheAspArgAsnTyrGlyLeuArgThrPheLeuPheCysT 1854
1838 :.....:.....:.....:.....:
1839 GATTTGAAGTTTTCAGGGTGAAGAGGTCAGGATTTGAATTTCTC 4344
1854 hrProPhe.....ThrProAspGlyArgAlaHisGlyLys 1865
1855 :.....:.....:.....:.....:
1856 GGCATATCCGAAAGGAGAGAAAAAACCCAGAC.....AATGAA 4382
1866 LeuProGluGlnHisLysArgLysThrLeuLeuSerThrAspHisAlaP 1882
1867 :.....:.....:.....:.....:
1868 TTTTGGCAATATGTGATGCGAGAGAACATATATCACTGATTAATTT 4432
1882 eProTyrIleLysThrArgIleArgValLysHisArgGlnIleThrVal 1899
1883 :.....:.....:.....:.....:
1884 ACCTGAAATTTTAAAGTGTGAGGTCAAGTCTGTTTCAATGTTGAAA 4442
1899 euThrProValGlnValAlaIleGluAspMetGlnLysLysThrArgGlu 1915
1899 :.....:.....:.....:.....:
1900 TCAGCCCTCGAGAAATGCCATCGAGACCATGACGAGCAAGCAAG 4532

```

```

1916 LeuAlaPheAlaThrGlnIleAspProAspAlaLysMet..... 1929
1917 :.....:.....:.....:.....:
1918 ATCAACACAGTGTGTACAGACACCTGGATGACCCACCTGCCCCAA 4582
1930 ....LeuGlnMetValLeuGlnGlySerValGlyProThrValAsnGln 1945
1931 :.....:.....:.....:.....:
1932 CCCGCTTCATGCTCTCGAAGCGATGTGAGCCAGCTTCATGGGGG 4632
1945 yProLeuGlnValAlaGlnValPheLeuAlaGlu.....IleProGlu 1959
1946 :.....:.....:.....:.....:
1947 GCTTGCACAACTACGAAAGCCCTTCTTACAGACCGGTACTCGAGAG 4682
1960 AspProLysLeuPheArgHisAsnLysLeuArgLeuCysPheLysAs 1976
1961 :.....:.....:.....:.....:
1962 CACCTGAGGCCCATGAAAGATGAGAGCTC.....AAGA 4720
1976 rPheCys..... 1978
1977 :.....:.....:.....:.....:
1978 CCGTTCACGAGATTCCTTTCTGGCCGAGGATCAGATCCATG 4770
1979 ..LysLysCysGlnAspAlaLeuArgLysAsnLysAlaLeuIleGlyPro 1994
1980 :.....:.....:.....:.....:
1981 GAGCAAAAGTCAGGAGGCTGAGG..... 4796
1995 AspGlnLysGluTyrHisArgGluLeuGlnLysArgAsnTyrCysArgLeuAr 2011
1996 :.....:.....:.....:.....:
1997 .....CGTTCACGAGATGAGGCTGTTCAACACGTGAA 4837
2011 gGluAlaLeuGln 2015
2012 :.....:.....:.....:.....:
2013 GAAAAAGGTGAG 4850

seq_name: /cgn2_6/prodata/2/lna/6A_COMB.seq:US-08-923-137-2
seq_documentation_block:
; Sequence 2, Application US/08923137
; Patent No. 6083716
; GENERAL INFORMATION:
; APPLICANT: Wilson, James M.
; APPLICANT: Farina, Steven F.
; APPLICANT: Fisher, Krishna J.
; TITLE OF INVENTION: Chimpanzee Adenovirus Vectors
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr., P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: United States of America
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923.137
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,700
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNPVN.021CIP1USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 36519 base pairs
; TYPE: nucleic acid

```

```

; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-08-923-137-2

```

```

Alignment_scores:
    Quality: 191.00      Length: 1635
    Ratio: 0.302        Gaps: 88
    Percent Similarity: 38.716    Percent Identity: 21.101

```

alignment\_block:

US-09-736-968a-2 x US-08-923-137-2/reverse ..

Align seg 1/1 to reverse of: US-08-923-137-2 from: 1 to: 36519

```

125 ArgArgTyrGlnTyrLeuSerAlaIatYserProValThrThrAspTh 141
    ||||| : : : : : ||| : : : ||| : :
8593 GGTGGCTACGCCAGTCTTCTCCGGGCGACGTACCAAGAACCTACCGCGC 8544
    : : : : : : : : : : : : : : : :
141 rGlnArgGlnArgGlnLysGlyLeuProArgGlnValPheGlnGlnAspA 158
    : : : : : : : : : : : : : : : :
8543 GCGGGCGGACGTGCCCCCTGCGGCCCT..... 8516
158 lAserGlyAspGlnArgSerGlyProGlnAspSerAsnAspSerArgArg 174
    ||||| : : : : : : : : : : : :
8515 ..... GCTCCGGGCCCGAGCGCCCTGCGCGCGCGCGC 8480
175 GlySerGlySerPro.....GlnAsp..... 181
    : : : : : : : : : : : : : : : :
8479 GGTCTCTGCGCGCTGCTTCAACATGGAAGAGCAGCCCGCCGTCGCC 8430
182 .....ThrProArgSerSerg 187
    ||||| : : : : : : : : : :
8429 CCAGCGCGCAAGAGAAGAGGAGGAGCGTGTACACCCAGGAGGACG 8380
187 lYAlaSerSerTlPheAspLeuArgAsnLeuAlaIaAspSerLeu... 202
    : : : : : : : : : : : : : : : :
8379 GGACCTCTGCAAGCCATCGACGTGCGCACCAATGGCGGTGAGATCAAG 8330
203 .....LeuProSerLeuLeuGln.....Ar 209
    ||||| : : : : : : : : : :
8329 TACCAATCTGGACCTCCCGCGCGCCCTGGAAAAAATCTGCAAGTCAACG 8280
209 gAlaAlaProGlnAspValAsp.....ArgArgAsnGlnThrLeuA 223
    ||||| : : : : : : : : : : : :
8279 GCGCGCGCGCGTGCCTGACTGACTCCGAGGCGTCTCAGAGACCTGG 8230
223 rGArgGlnHisArgProProAlaLeuLeuThrLeuTyrProAla..... 237
    : : : : : : : : : : : : : : : :
8229 ACAGCTCGGCGCTGCGCGCCCTGCTCTGCTCCGACCGCGTCGCGCC 8180
237 ..... 237
8179 GAGGTCTGACT3CCTCCCGCGGCGGCTCGTACAGATGACACCATCGA 8130
238 .....ProAspGlnAspGlnAlaValAlaArgCysSera 249
    ||||| : : : : : : : : : : : :
8129 GGCAGAGAGGCGCAGGCGGACACACAGACGTAGTACAGCAC.....G 8086
249 rGProGlnProArgGlnHisPheGlyGlnArgIleLeuValLysCys 265
    : : : : : : : : : : : : : : : :
8085 AGATGCAAGCGCCACGACTCCACTTCCACTGAATTCCTCGTCAAGGA 8036
266 LeuSerLeuLysPheGlnIleGlnIleGlnProIle..... 277
    : : : : : : : : : : : : : : : :
8035 ACTCAGGTACACTCGTCGACAGACGTCCCGTCAGCGCTTGAGACA 7986
278 .PheGlyIleLeuAlaLeuTyrAspValArgGlnLysLysIleSer. 293
    ||||| : : : : : : : : : : : :
7985 CTGTGCGCCCTTG.....TATTAACACAAGCAGAGTGTTCGG 7948
294 .....GlnAsnPheTyrPhe...AspLeuAsnSeraSperMetLys 306

```

```

7947 CGCGCGGTGCGGATTTCTTACTTCATCATCATCAACAGCATTCCTCAT 7898
    : : : : : : : : : : : : : : : :
307 GlyLeuLeuArgAlaHis.....GlyThrHisProAlaIleSe 319
    : : : : : : : : : : : : : : : :
7897 TGGTGGCAGAGATCCAGTTCTTCCCATCGGCTCGCATCTCGACCGCA 7848
319 rThrLeuAlaArgSeraIaIlePheSeraValThrTyrProSerProAspI 336
    : : : : : : : : : : : : : : : :
7847 AAGCTC..... 7841
336 lPheLeuValIleLysLeuGlnLysValLeuGlnGlnLysPleSer 352
    ||||| : : : : : : : : : : : :
7840 ..TTCTCACTACATGATGTGAAACCTACACCTGATGGGTCCTTCGCG 7793
353 GLucsCysGlnProTyrMetValLeuLysGlnValAspThrAlaLysAs 369
    : : : : : : : : : : : : : : : :
7792 AACGACGTGCTCCCTTCATGCTGATCAAGAACTC.....TCGG 7752
369 nLysGlnLysLeuGlnLysLeuArgLeuAlaIaGlnGlnPheCysThra 386
    : : : : : : : : : : : : : : : :
7751 GACAGACCGGCTGTCGAGCTCGCCCTC..... 7724
386 rGLeuGlyArgTyrArgMetProPheAlaThrPheAlaValHisLeuAla 402
    : : : : : : : : : : : : : : : :
7723 .....GATCTGCCCTCCACCTCAATGGATGCTGCGCAC..... 7688
403 AsnIleValSerSeraIaGlnLeuAspArgAspSeraSperSergLugl 419
7688 ..... 7688
419 yGlnArgArgProAlaTrpThrAspArgArgArgGlyProGlnAspA 436
7688 ..... 7688
436 rGAlaSerSerGlyAspAspAlaCysSerPheSerGlyPheArgProAla 452
    ||||| : : : : : : : : : :
7687 .....GGGAGACCCCGCACCTTCTACGCGCGACACCCCGAA 7652
453 ThrLeuThrValThrAsnPheLysGlnGlnAlaGlnArgLeuSera 469
    : : : : : : : : : : : : : : : :
7651 AATATGCGCTGCGGCGAGCATTCCTGCTACGAGGACCGCTCGCAAC 7602
469 pGlu.....AspLeuPheLysPheLeuAlaAspMetArgArgPros 483
    ||||| : : : : : : : : : :
7601 CGCCCTGCGCGCTGATCTCTG.....A 7579
483 eSerLeuLeuArgArgLeuArgProValThrAlaGlnLeuLysIleAsp 499
    : : : : : : : : : : : : : : : :
7578 CTTCCTCTCTCGC..... 7565
500 lLeSerProAlaProGlnAsnProHisPheCys..... 510
7564 .....GCCAACCCCATCTCGCGACTGGGCGCTCGAGCA 7530
511 .....LeuSerProGlnLeuLeu.....HisIleLysProI 521
7529 GCACGGGCTCAGCGACCCGAGAGCTCACCTACGAGAGCTTAAGAAC 7480
521 yrProAspProArgGlyArgProThrLysGlnIleLeuGlnIlePheProAla 537
    ||||| : : : : : : : : : : : :
7479 TGCCCAACGTCAAGGCGCTCG.....CGATTCTGGAA..... 7445
538 ArgGlnValTyrAlaProHisThrSerTyrArgAsnLeuLeuTyrValTy 554
    ||||| : : : : : : : : : : : :
7444 .....CTCTACATCGT 7434
554 rProHisSerLeuAsn.PheSerSeraArgGlnLysSer..... 566
7433 GGGCACAATCAACGCGTTTCGACGAGATCGTCTGCGCCAGAGTCA 7384
567 .....ValArgAsnLeuAlaValArgVal.GlnTyrM 577
    ||||| : : : : : : : : : : : :

```

```
7383 TCACACACCGGCGGAGGTGCCCCAACCTTTCCGCATCAGTCGCAACTTT 7334
577 eUthr.....GlyIuAsProSerGln..... 584
:::|||||:::
7333 ATGCCCGCGGGGCGCAAGATCTCTTCAACGATGTACTTTCCGTTGCC 7284
585 .....AlaLeuProValIlePheGlyLysSe 593
:::|||||:::
7283 CAACCCCGCCTACAGAAAGCCACGCACTTCCAGCTGTG...GGAGCAGG 7237
593 rSerCysSerGluPheThrArgGluAlaPheThrProValIleYrHisA 610
:::|||||:::
7236 GGGCTGGGAGGATATCGACTTTCAGACA.....CCAGTCTCAAGGTC 7193
610 snLysSerProGluPheTyGluGluPheLysLeuHisLeuProAlaCys 626
:::|||||:::
7192 ATGTCAGGAGACACTTTCGCCCTGC.....GCACACTCCCGCG... 7152
627 ValThrGluAsnHisLeuLeuPheThrPheTyrHisValSerCysG1 643
7151 .....CA 7150
643 nProArgProGlyThrAlaLeuGluThrProValGlyPheThrTrpIleP 660
|||||
7149 AGGCGCGCCAGGC.....CTACCCCTCCCGCGGAGAGGCGCTGTC 7106
660 rOleu.....LeuGlnHisGlyArgLeuArgThrGlyPro 671
|||||
7105 GCGTCACAGCGCGTCACACAGTCTACATGCTAGAGCTC..... 7068
672 PheCysLeuProValSerValAspGlnProProSerTySerValLe 688
|||||
7067 .....TTACCGTCAGATCAGAGCGGGGTTCCTCCCTCAAGA..... 7032
688 uThrProAspValAlaLeuProGluMetArgTrpValAspGlyHisLysG 705
:::|||||
7031 .....GTACTGGAGGAGCGCGGAGAGTTCCTTCACACCGC 6995
705 LyValPheSerValGluLeuThrAlaValSerValHisProGlnAsp 721
:::|||||
6994 GAGCTGTGAAACAAAGGAGCACTCAAGTATGACATCCACGAGAC 6945
722 ProTyLeuAspLysPhePheThrLeuValHisValLeuGluGluGlyAl 738
|||||
6944 CCTGAGACTA.....CTGCGCTCTGGACG 6922
738 aPheProPheArgLeuLysAspThrValLeuSerGluLysValGluG 755
|||
6921 TCCTAGTCACCGC..... 6909
755 InGluLeuArgAlaSerLeuAlaAlaLeuArgLeu..... 766
|||||
6908 .....CGAGCTCGTCCCAAGCTGCAGACTCTACGCCCACTTCAT 6867
767 .....AlaSerProGluProLeu.....ValAlaPheSe 776
:::|||||
6866 CCGGAGACTCGGTGGAGCTGCCACGCGCCACTTCACATCTTCCAAGGC 6817
776 rHisHisValLeu.....AspLysLeuValA 785
:::|||||
6816 CCACCACTCTCTCCAACTGCGCATGTCATTTCCGCCAGATCGCTA.... 6771
785 rGluLeuValIleArgProProIleIleSerGlyGlnIleValAsnLeuGly 801
|||
6770 .....CCGCGCGGAAAGCCGAGTCGACCAACCTGGG 6738
802 ArgGlyAlaPheGluAlaMetAlaHisValValSerLeuValHisArgSe 818
|||||
6737 CCGGGGCTCTCGCCCTCGCAGAGCTA.....TG 6703
818 rLeuGluAlaAlaGlnAspAlaArgGlyHisCysProGlnLeuAlaAla 835
|||||
6702 ACTACGTGGCGCGCCAGCATCGCGGGG.....GCGCTGTACATCCC 6662
835 yValHis..TyrAlaPheArgLeuProGlyThrGlu..... 846
:::|||||
6661 ACCTACATCGGCATTCCTCGAGAGCCCTCTATACAGATCTGCGG 6612
847 .....ProSerLeuProAspGlyAlaProProValThrValGln 859
|||
6611 CATGTACGGCTCGCGCTTACCATCCCATGCCCTGGGGCAGCGCCCTCA 6562
860 AlaAlaThrLeuAlaArg.GlySerGlyArgProAlaSer..LeuTyGlu 875
:::|||||
6561 GTCCCTACGAGCGCGCTGCGCGCAATGGCAACTGCCCGCGAGC 6512
876 AlaArgSerLysSerLieserSerAsnProAspLeuAlaValAlaArg 892
:::|||||
6511 GATCTGGCCACTTCATCACTAGCTTTCAGCCCGGACCTCTCCCGCAT 6462
892 oGlySerValAsp...AspGluValSerArgIleLeuAlaSerLysLeuL 908
:::|||||
6461 CTTCACCATCGATGCGCGCCCGCGAGAGGTGATGTCGACCCCTG. 6413
908 euHisGluGluLeuAlaLeuGlnTrpValValSerSerValValArg 924
6413 ..... 6413
925 GluAlaIleLeuGlnHisAlaTrpPhePheGlnLeuMetValLysSe 941
6413 ..... 6413
941 rMetAlaLeuHisLeuLeuLeuGlyGlnArgLeuAspThrProArgLysL 958
|||
6412 .....CCCCCTCTGCT 6400
958 euArgPheProGlyArgPheLeuAspArgIleThrAlaLeuValLysSer 974
|||
6399 CCGGCAAGGGCGGAGCGCTCTGTAATGAGCCCTGCGCGGAG 6350
975 ValGlyLeuGluValIleThrArgValHisLysAspValGluLeuAlaG1 991
|||||
6349 GTGGCCACCACTGTG..... 6335
991 uHisLeuAsnAlaSerLeuAlaPhePheLeuSerAspLeuLeuSerLey 1008
|||||
6334 .....GACCTCATACCCCTGC 6319
1008 aLAspArgGlyPheValPheSerLeuValArgAlaHisTyLysGluVal 1024
:::|||||
6318 ATAATCGGGGCTGCGAGGTCAGAAATCGTC..... 6290
1025 AlaThrArgLeuGlnSerSerProAsnProAlaAlaLeuLeuThrLeuArg 1041
|||
6289 .....CCGACGAGCTGACACCGCTC... 6269
1041 gMetLurPheThrArgIleLeuCysSerHisGlnHisTyValThrLeuA 1058
|||
6268 .....TTCCCGAATGAATGCTGTCGCGCGAGTACGTGCACTCA 6226
1058 snLeu.Pro.....CysCy 1062
|||
6225 ACATCGCGCGCAAGAGCGCGCCAGCAAGAAACCAACCATCGCGC 6176
1062 sProLeuSerProProAlaSerProSerProSerVal..... 1074
|||
6175 TCCA...TCGCGAAGCTCTCTCCAAAGCCCTTACGCTCTTGGCCAC 6129
1075 ..SerSerThrThrSerGlnSerSer.....ThrPheSer 1085
|||||
6128 CAAGCTCGACACAAAGATGCTGTTTTCGACCAATGACAGGAGGCG 6079
1086 SerGlnAlaProAspProLysValThrSerMetPheGluLeuSerGlyPr 1102
|||
6078 TCTTGAAAGGATCTCGCGCGGACCGCTCA.....ATATCAAAATCC 6038
```

```

1102 opehAygGlnGlnHisPheLeuAlaGlyLeuLeuThrGluLeuAla 1119
      |||
      |||
6037 TCCTCGTTT.....CTAGAACTGACAACTGAGTG 6007
1119 euAlaLeuGluProGluAlaGluGlyAlaPheLeuLeuHisLysLysAla 1135
      |||
      |||
6006 CCGAGGTCATGCCGCTTCGAGAGGGGAATCCTACCCCAACAGCTGGCG 5957
1136 IleSerAlaValHisSerLeuLeuCysGlyHisAspThrAspProArgTy 1152
      |||
      |||
5956 CTCTCG.....GACAGCATCCGGAGGA 5934
1152 rAlaGlu.....AlaThry 1157
      |||
      |||
5933 CAGTAGAGAGAGAGAGGACCGCACCTTTTATACCCCCCGCGCGGA 5884
1157 alLysAlaArgValAlaGluLeuTyLeuProLeuSerIleAlaArg 1173
      |||
      |||
5883 CCCCCGCTACGTGGCTTACACTTACAGCCAAATCACCTTCTGGATGTG 5834
1174 Asp.....ThrLeuProArgLeuHisAspPheAl 1183
      |||
      |||
5833 GACAGAGGGGACATGTGTTGCATACCTCGAAAGGTGAGACCGGTGT 5784
1183 aGluGlyProGlyGlnArgSerArgLeuAlaSerMetLeu..... 1196
      |||
      |||
5783 GGACAAAGACCGCTACCCGTCCACGTGGCTCTTGTGTGGCTGGA 5734
1197 .....AspSer 1198
      |||
      |||
5733 CCGCGGCTTCGTCTCGAGTGGCGGGGTTCTCTACGAGGAGACCGC 5684
1199 AspThrGluGlyGlu.....GlyAspIle.A 1207
      |||
      |||
5683 GGCACACCGCTCGAGACCGGCCATTAAGTGTGTACGGGACACGGA 5634
1207 laGlyThrIleAsnProSer..... 1213
      |||
      |||
5633 CAGCTCTTGTGTACCAAGCGGGACACGACTGATGAGACCAAGTA 5584
1214 .....ValAlaMetAlaIleAlaGly.....ProLeuAl 1224
      |||
      |||
5583 AGAAAGCATCAAAAGAACGGCGGAACGTGTTTGAACCGGACAG 5534
1224 aProGlySerArgLaserIleSerGlnGlyPro.....ProThrLas 1239
      |||
      |||
5533 CCGGACCTCACCTGTCTGTGAGTGGAGACCTGTGCGCCACTGCGG 5484
1239 eATrGAlaGlyCysAlaLeuSerAlaGluSerArgThrLeuLeuAla 1255
      |||
      |||
5483 CGCGGAC.....GCCTAGCCCCCGAGTGC.....GTCTTCTGTGCC 5447
1256 CysValLeuTrpValLeuLysAsnThrGlnProAlaLeuGlnArgTy 1272
      |||
      |||
5446 CCAAGCTCTACGCCCTCAAGTCC..... 5433
1272 pAlaThrAspLeuThrLeuProGlnLeuGlyArgLeu..... 1284
      |||
      |||
5422 .....CTCCTCTGTCTCCGCTGGGAGACTTCCAAAGTAAGC 5383
1285 .....LeuAspLeuLeuTy 1289
      |||
      |||
5382 TCCGCGCCAAAGGCCACGCGCCGAGCGCTCAACTGCAACTATGATGTC 5333
1290 LeuCysLeuAlaAlaPheGluTyLysGlyLysAlaPheGluArgI 1306
      |||
      |||
5332 AATGCTACTCGGCC...GACGGCAGGGCGCATGG...GAGCGGTT 5289
1306 eAsnSerLeuThrPheLysLysSerLeuAspMetLysAlaArgLeuG 1333
      |||
      |||
5288 TTCGAC...AGCCGATGAGCTGAAGCGCACCTCCCA 5251
1323 lAlaIleLeuGly..... 1327

```

```

5250 GCGCCAGCCCGCGGACGCCCTTCACCTGTGACGAGACCACTGACG 5201
      |||
      |||
1328 .....ThrIleGlyAla.....ArgI 1333
5200 CCGAGCTTCGACCTCGAAGAGACCGGACCGCTGCGCGTGAAGCCCA 5151
1333 nGluMetValAlaArgArgSerArgGluArgSerProPheGlnAsnProGluA 1350
      |||
      |||
5150 TCGTGTGGTCCCTTACTCCGCGAGCGTCC.....AACCGCGCA 5110
1350 sn.....ValArgTrp..ArgLysSerValThrHisTrpLysGlnThr 1364
      |||
      |||
5109 ACAGAGAGGTGTCTGTGATCGATGCGATGAGACACATCACCAAGCTCT 5060
1364 eAspArgValAlaLysThrLysAspGluMetGlnHisGluAlaLeuVal 1380
      |||
      |||
5059 GGGACCGCTGAGACTCTTGCACAAACCCCTCCAAATGGCCATGGCC 5010
1381 GluGly.....AsnLeuAlaThrGluAlaSerLeuVal 1392
      |||
      |||
5009 GACGCACTCAACCGCTGAAAACTTGGCTGCTGCCAGAGACTCTATC 4960
1392 lLeuAspThrLeuGluIleIleValGlnThrVal..... 1403
4959 CTTGGGGGAGAGCGCTCTTGGCGAACTGTGCGCGGAGAACATGACAG 4910
1404 .....MetLeuSerGluAla.....ArgGluSer..... 1411
      |||
      |||
4909 TGGCGAGATGATGAACAGTGTGGCCCCCTCTCCGGAGAGACGAGC 4860
1412 .....ValLeuGlyAlaValLeuLysValValLeuTy 1422
      |||
      |||
4859 TGCTCTCCCTCAACTACCACTGACGCGGTATCGGGGTATCTACGG 4810
1422 rSerLeuGlySerAlaGlnSerAlaLeuPheLeuGlnHisGlyLeuAlaT 1439
      |||
      |||
4809 CCGCACCGGCTGGGCAAGTCCAGCTG...CTCCGAACTGTCTTGG 4763
1439 hGlnArgAlaLeuValSerLysPheProGluLeuLeuPhe 1452
      |||
      |||
4762 CCCAG.....CTCATCTCCCCCGCCCGAAACCGTGTTC 4728

seq_name: /cgn2_6/plodata/2/lna/5B_COMB.seq:us-08-804-227C-7
seq_documentation_block:
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.

```

```
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ. ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-09-804-227C-7
```

```
alignment_scores:
Quality: 182.00      Length: 1649
Ratio: 0.285        Gaps: 89
Percent Similarity: 38.690      Percent Identity: 21.953
```

```
alignment_block:
US-09-736-968A-2 x US-08-804-227C-7 ..
```

```
Align seg 1/1 to: US-08-804-227C-7 from: 1 to: 44377
```

```
2 AlaAlaSerGluArgAlaPheAla.HisLysIleAsnArgThrValA 18
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23189 GCGTGTCTGACGAGATCAACGCGACCGCCGTCACCGCGTC 23238
18 LaAlaGluValArgLysGlnValSerArgGluArgSerGlySerProHis 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23239 TCCGT.....GCTCGACTGTGCACCGACGCGCGACCGCCGCC 23282
35 SerSerArgCysSerSerSerLeuGlyValProLeuThrGluValA 51
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23283 CGTGGCCGTGC..... 23294
51 IGIuProLeuAspPheGluAspValLeuLeuSerArgProProAspAlaG 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23295 ..... 23316
68 IuProGlyProLeuArgAspLeuValGluPheProAlaAspAspLeuGlu 84
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23317 GTCCAGGCCCTGGCCGA..... 23333
85 LeuLeuLeuGluProArgGluCysArgThrThrGluProGlyTyleProly 101
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23334 .....CCTCGGCGACGACCGCCACCGCT..... 23354
101 sAspGluLysLeuAspAlaGlnValArgAlaValGluMetTyrTleG 118
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23355 .....GTGACCGTGTACCTCGCGCGGT.....CGCCACCGCC 23388
118 IuAspTyrValIleValHisArgArg..... 126
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23389 CCCGAGAACTGCGTGCACCGCGGTGCCACGCTGTGGGGCTGGGGCCG 23438
127 .....TyrGlnTyrLe 130
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
23439 GGTGGCCGCGGTGACCTGCCAGAGTGTGGGGCGGCGTCATCGACTTC 23488
130 uSerAlaAlaIaTyrSerProValThrThrAspThrGlnArgGlnL 147
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23489 CCGCGGCGCCGACCGCGGTCTGGACCGCGGTCTGGCGCGCGCT..... 23531
147 ysgLysLeuProArgGluValPheGluGlnAspAlaSerGlyLyspGluArg 163
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23532 .....CCTCGCCCAACC.....CGCGCG 23549
164 SerGlyProGluAspSerAsnAspSerArgArg.....Glyse 176
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23550 CGAGGACAGATCGCCGTACGATGCGGGCGCTGTGGCGCGCGGTCC 23599
176 rGlySerProGluAspThrProArgSerSerGlyAlaSerSerIlePheA 193
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23600 TGGGGAACTCCGCGCGACTCCCGCCCC..... 23627
193 sPLeuArgAsnLeuAlaAlaAspSerLeuLeuProSerLeuLeuGluArg 209
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23628 .....GGCTGGCGCGCGCGCGGCGACCGTCT.....CATGCG 23660
210 AlaAlaProGluAsp.....ValAspA 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23661 CGGCGACCTACAGAGGTGCCCGCGCGACTGTCCGTCCCTCGAGG 23710
215 ..... 23711
23711 ACGCGCGGACCGCGTGTGCTGGCGCGACCGCGCGCGCGCGCGAGCC 23760
217 rGArgAsnGluThrLeuArgArgGlnHisArgProProAlaLeuLeuThr 233
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23761 GCGCGCGCGGACTGACCGCGGTCTCCGTCCGTCGCCGCGCGCTGGAGCT 23810
234 LeuTyrPro.....AlaProAspGluAspGluAlaValGlu 245
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23811 CACGACCGCGCGCGACTGGCGCGCTGTCTGACGACGACGCGCGCACCG 23860
245 uArgCysSerArgProGluPro..... 252
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23861 TCGCGGTGACGCGCGCGCGCTGTGCGCGCGCGCGCGCGCGCGAGCG 23910
253 ..ProArgGluHisPheGlyGlnArg.IleLeuValLysCysLeuSerLe 268
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23911 GCACCGCGCGGACAT...CGCGCGCGCTCGCGCGCGCGCGCGCGCG 23957
268 uLysPheGluIleGluIleGluProIlePheGlyTyleLeuAlaLeuTyrA 285
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23958 CGGCACTGTGTGACTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG... 24003
285 sPValArgGluLysLysIleSerGluAsnPheTyrPheAspLeuAsn 301
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24004 .....GTGCTGTCTCTCGTGTGCTCGGAGTGTGG 24033
302 SerAspSerMetLysGlyLeuLeuArgAlaHisGlyThrHis..ProAl 317
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24034 GCGCGCGCGCGCGCGCGCGCGCTACGCGCGCGCGCGCGCGCGCGCTCGAGCG 24083
317 aIleSerThrLeuAlaArgSerAla.....IlePheSerValThr 331
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24084 GCTGGCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCTCGTGGCT 24133
331 yRPro..... 332
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24134 GGAGCGCGCTGGCGCGGAGGACGCGCGCGCGCGAGTGCAGGCGGAGTTC 24183
333 .....SerProAspIlePheLeuVal11 340
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24184 CTCAGCGCGCGCGCGCGCTGCGTCCCTCGACCGCGCGAC...CAGCGGCGCG 24230
340 eLysLeuGluLysValLeuGlnGlnGlyAspIleSerGluCys..Cys.. 355
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24231 GACCTTGCGCGCGCATGCTGAGCGCGGC.....AGCGCTCGGTGCGG 24274
```



```

356 .....GluProTyrMetValLeuLysGluValAspThr 367
      |||||
24275 TCGGCAGCTGAGTGGACGGTTGCGCCCTCTACACCTGGTGGCT 24324
      |||||
367 LAlYsAnLysGLuLysLeuGLuLysLeuAlaIaGLuInPhe 383
      |||||
24325 CCCCCCTACTCTTCGACGACATCCCGGACGTGACGGCTGGCGGCG 24374
      |||||
384 CysThrArgLeuGLYArgTyrArgMetProPheLatrPThrAlaValH1 400
      |||||
24375 CGAACTGCCCCGACGACCGAGACTGAC.....CA 24406
      |||||
400 sLeuAlaAsnLLeValSerSerAlaGLYInLeuAspArgAspSerAsps 417
      |||||
24407 CCTC.....CGAACTGTCGGCGAGCTGACGGCGAGTCCGGC 24444
      |||||
417 eRGLuGLYGLUArgArgProAlaTrPThrAspArgArgGLYPro 433
      |||||
24445 CACAAGGGGCGACCCACTGCTGCGGTGGCGGACACAGCCGCGCG 24494
      |||||
434 GlnAspArgAlaSerSerGLYAspAlaCysSerPheSerGLYPheA 450
      |||||
24495 CGTCTCGGACAGCTCTCGCGGACGGGTGAGCAGCGCCGCGCTTCC 24544
      |||||
450 rGPProAla.....ThrLeuThrValThrAsnPheLysGLuGLu 463
      |||||
24545 GCGACCTGCGCTGACTCGCTGACCGCCCTCGAACTG..... 24582
      |||||
464 AlaGLUArgLeuSerAspGLUAspLeuPheLys.....PheLe 476
      |||||
24583 CCGCAGCGGCTGACGACCGGCGCTCAACTGCCACCTCCCTGGGT 24632
      |||||
476 uAlaAspMetArgArgProSerSerLeuLeuArgArgLeu..... 489
      |||||
24633 CTTCGACCACTCCAGCCGGCGGCGTGGCGGACACTCGGTGAGGAC 24682
      |||||
490 .....ArgProValThrAlaGLuLeuLys 497
      |||||
24683 TCCTGGCGGGAAGACACCGCCGCGGCGGCGCCGACACCCGGTA 24732
      |||||
497 sILeAspLeuSerProAlaProGLUAsn.....ProH 508
      |||||
24733 CGGACGAGGAGCCCATGCCATGCGCATGGCTGCGGTCGCGCGCG 24782
      |||||
508 iSPheCysLeuSerProGLULeuLeu.....HisILeLysPro 520
      |||||
24783 CGGGGTGAGTCCCGGAGACCTGTGGACCTGTCAGCCGTGGGACCG 24832
      |||||
521 TyrProAspProAlaArgGLYArgProThrLysGLuILeGLuPheProAl 537
      |||||
24833 ACGCATATACCCCTTCCCGAACACCGGAGTGGACACAGACCCCTC 24882
      |||||
537 aArgGLU.....ValTyrAlaProHisThrSerTyrArgAsn 549
      |||||
24883 TACGACCCCGGACCCGACTGCGCGGACACCACTACGTGCGGAGGG 24932
      |||||
550 LeuLeuTyrValTyrProHisSerLeuAsnPheSerSerArgGLNLYSe 566
      |||||
24933 .....CGGGTCTCTGCACAGCGGGCGG 24955
      |||||
566 rValArgAsnLeuAlaValArgValGLN.....TyrMetThrGLYGLu 580
      |||||
24956 AGTTGACACCCCGCTTCTTCGCGATACAGCCCGGAGGCCCTGCGCATG 25005
      |||||
581 AspProSerGLNAlaLeuProValILePheGLYLysSerCysSerGL 597
      |||||
25006 GACCCGACACAGCGGCTGATCTCTG.....GAGAGCTCTGGGAGGTC 25046
      |||||
597 uPheThrArgGLUAlaPheThrProValValTyrHisAnLysSerProG 614
      |||||
25047 CTTCGAACGGCGCGCATGACCCGCTCAACTGCGCGGACAC..... 25089

```

```

614 LuPheTyrGLuGLuPheLysLeuHisLeuProAlaCysValThrGLuAsn 630
      |||||
25089 ..... 25089
      |||||
631 HisHisLeuLeuPheThrPheTyrHisValSerCysGLNProArgProGL 647
      |||||
25090 .....CGCACCGG 25097
      |||||
647 yThrAlaLeuGLUThrProValGLYpHeThrTrPLeProLeuLeuGLN 663
      |||||
25098 GGTCTGTCGCGACCAAC...GGACAGCACTACGTGCGCTCTCCAGG 25144
      |||||
664 .....HisGLYArgLeuArgThrGLYProPheCysLeu 674
      |||||
25145 ACGGCGACGAGAACTTGACGGCTATCATGCCACCGGC..... 25182
      |||||
675 ProValSerValAspGLNProProProSerTyrSerValLeuThrProAs 691
      |||||
25183 .....AACTCGGACAGGTATGTCGGCGG 25208
      |||||
691 rValAlaLeuProGLYMetArgTrPValAspGLYHisLysGLYValPheS 708
      |||||
25209 GCTCTCTAC.....GTCCTTG 25225
      |||||
708 eValGLuLeuThrAlaValSerSerValHisProGLNAspProTyrLeu 724
      |||||
25226 GACTGGAGGGCCCCCGCGTAC..... 25248
      |||||
725 AspLysPhePheThrLeuValHisValLeuGLuGLuAlaPheProH 741
      |||||
25248 ..... 25248
      |||||
741 eArgLeuLysAspThrValLeuSerGLuLysnValGLuGLuInLeuAla 758
      |||||
25249 .....GTGCACACCGCTGCT 25264
      |||||
758 rGLAserLeuAlaAlaLeuArgLeuAla..... 767
      |||||
25265 CGGCTCTCCGCGGACACTGCTGCGGTGACGTGACCTGCGCGCGCG 25314
      |||||
768 .....SerProGL 770
      |||||
25315 GAATGCGACTACCCCTCGCGGCGGCGGACAGGTATGTCACACCCGA 25364
      |||||
770 uProLeuValAlaPheSerHisHisValLeuAspLysLeuValArgLeu 787
      |||||
25365 GATGCTGTGTGAGTGTGCCCT.....CAAGCGAG 25393
      |||||
787 alLeuArgProProLLeuLeuSerGLYInLLeValAsnLeuGLYArgGLY 803
      |||||
25394 CGGTGTCGCGGAGCGCGGACAGGCGTTGCGGAGGCGGCGGACGGG 25443
      |||||
804 AlaPheGLUAlaMetAlaHisValAlaSerLeuValHisArgSerLeuGL 820
      |||||
25444 GTGCTGTGCGCGAGGAGCGGGATGCTGTGTGAGCGGTGCGGA 25493
      |||||
820 uAlaAlaGLNAspAlaArgGLYHisCysProGLNLeuAlaAlaTyrValH 837
      |||||
25494 GCGCGACAG.....AAGGCCAT...CGGTATGCGCGGTGTGCGGG 25534
      |||||
837 iSTyRAlaPheArgLeuProGLY.....ThrGLUProSer 848
      |||||
25535 GCACTGCGCTTACACGAGAGGTGCGGACAGAGGCTCACCGACCCAGC 25584
      |||||
849 LeuPro.....AspGLYAl 853
      |||||
25585 GGGCGCGCCACGAGCGGTGATACGGAGCGGTGCGGACGAGGGGCT 25634
      |||||
853 aProProValThrValGLNAlaAlaThrLeuAlaArgGLYSerGLYArgP 870
      |||||
25635 GACCGCGCGAGCTGTGAGCGGTGAG...GCGCAGGACCGGCGACGC 25681
      |||||
870 ro..... 870

```

```
25662 CGCTGGGAGACCCATCGAGCGCGCGCTGTCGCCACCTACGCGCG 25731
871 .....AlaSerLeuTyrLeuAlaArgSerIleSerSe 883
25732 GACCGGGCGAGACGCCCTGTGGTGGTTCCGTGAG..... 25770
883 rSerAsnProAspLeuAlaValAlaProGlySerValAspAspLys 900
25771 .TCGAACATCGCGGCACACCGCGCGCGGTGGCGGTGATCA 25819
900 eArgGlyLeuAlaSerLysLeuLeuHisGluGluLeuAlaLeuGlnTyr 916
25820 AGATGGTCTGGCG.....CTGGCCACGCGAGCTGGCGGACCCCTG 25863
917 ValValSerSerSerAlaValArg.....GluAl 926
25864 CACCGCTGACGCGCGTGTCCAGATGATGGAGCGCGCGCGCGTGA 25913
926 aIleLeuGlnHisAla.....TyrPhePheGlnLeuMetValLys 941
25914 GTTCTGTGACGAGGCGACGCGCTGTCTC..... 25941
941 eMetAlaLeuHisLeuLeuGlnGlyGlnArgLeuAspThrProArgLys 957
25942 .....CAGCGGCGCGAGGCGCGCGCGG 25965
958 LeuArgPheProGlyArgPheLeuAspAspIleThrAlaLeuValGlySe 974
25966 .....GCGGCAATCTCTC 25979
974 rValGlyLeuGluValIleThrArgValHisLysAspValGluLeuAlaG 991
25980 GTTGGGATTCAGCGCG...ACCAACGCGCAGCTGTATCATGACGAGCGC 26026
991 LuHisLeuAsnAlaSerLeuAlaPhePheLeuSerAspLeuSerLeu 1007
26027 CGGAGCCACCGCG.....CCGCACTGCTC..... 26052
1008 ValAspArgGlyPheValIlePheSerLeuValArgAlaHisTyrLysGlnVa 1024
26052 ..... 26052
1024 AlaThrArgLeuGlnSerSerProAsnProAlaAlaLeuLeuThrLeu 1040
26053 .....GCGCCACACCGCGCGCGCGAGCGGAGCGTCT 26083
1041 .....ArgMetGluPheThrArgGlyLeuGlySerHisGlnHisTyrVa 1055
26084 GGTCCGAGAGTGTGGACAGAGTGACCGTCCCTGATGATGTCGCG 26133
1055 LThrLeuAsnLeuProGlySer.....ProLeuSerProPro 1068
26134 CACAGAGAACCGCGCGCTGCGGACAGCGCGCGCGCTGCGCGCACCT 26183
1068 LaSerProSerProSerValSerSerThrThrSerGlnSerSerThrPhe 1084
26184 GCTGGCGACCGCGGAGCTGACCGCGCGAGCGTGGCTCA..... 26224
1085 SerSerGlnAlaProAspProLysValIleThrSerMetPheGlnLeuSerGln 1101
26225 CCTCATCTACACCGCGCGCG..... 26247
1101 yProPheArgGlnGlnHisPheLeuAlaGlyLeuLeuLeuThrGlnLeuAla 1118
26248 ....TTCGACACGCGCGCGCGCTGTCTGGGAGAACTTCACGAGGCTGA 26293
1118 LaLeuAlaLeuGluProGluAlaGluGlyAlaPheLeuLeuHisLysLys 1134
26294 TCGCGCGCGCTCGACGACCTGTGAAAGC..... 26322
1135 AlaIleSerAlaValHisSerLeuLeu.Cys.GlyHisAspThrAspPro 1150
```

---

```
26323 .....CGACCGACCGCGCTGTCTGTGGGGGACACCGCGCGACCTC 26365
1151 ArgTyrAlaGluAlaThrValLysAlaArgValAlaGluLeuTyrLeuPr 1167
26364 CGA.....CCAGTCTGTGT...CGTCTTCCCGCGC 26392
1167 oLeuLeuSerIleAlaArgAspThrLeuProArgLeuHisAspPheAlaG 1184
26393 AGGCGTCCGAGTGGCGCGGAGATGGCCGACG.....GCTG 26427
1184 LuGlyProGlyGlnArgSerArgLeuAlaSerMetLeuAspSerAspThr 1200
26428 CTGGCCGCTCCAGCGGCTCCGAGCTC.....CTT 26456
1201 GluGlyGluGlyAspIleAlaGlyThrIleAsnProSerValAlaMetAl 1217
26457 CTTGGAGACCGCGCGCGCTGGACCTCGCGCTCGGCGCCACCTCGGCT 26506
1217 aIleAlaGlyGlyProLeuAlaProGlySerArgAlaSerIleSerGln 1234
26507 GGTCCGCTCTGGACGACTGCGCGCGGAAACCGCGCGCGCTCTCTCA. 26555
1234 LysProThrAlaSerArgAlaGlyCysAlaLeuSerAlaGlnSerSer 1250
26556 .....CCGGTCCAGCTGTGACGCGCGCTGTCTTCCATCATAT 26594
1251 ArgThrLeuLeuAla.CysValLeuThrValLeuLysAsnThrGluPro 1267
26595 GGT.....CTGCTCGCGCGACAGCTGGCGTTGCTGGGCGCGACGCG 26638
1267 LaLeuLeu.....GlnArgTyrAla 1273
26639 CCGCGTCTCGTCCACTCCACGAGCGAGATGCGCGCGCTACGTCTCC 26688
1274 ThrAspLeuThrLeuProGlnLeuGlyArgLeuAspLeu..... 1287
26689 GCGCGCTGACGCTGTGACAGCGCGCGCATCTGTGCGCGCGACGCA 26738
1288 .LeuTyrLeuCysLeuAlaAlaPheGluTyrLysGlyLysAlaPheG 1304
26739 GCGTGGCTGGCGCGTGGCC.....GCCAAGGCGGCAATGG 26773
1304 LuArgIleAsnSerLeuThrPheLysLysSerLeuAspMetLysAlaArg 1320
26774 TCGCGGTAC.....CTGTCCGAACGAGACTGTCGCCCG 26811
1321 LeuGluGlu..... 1323
26812 CTGGAGCCCTGGAGGACCGCGCTGCGCGCGCGCGCTCAACGCGCGGA 26861
1324 .....AlaIleLeuGlyThrIleGlyAlaArgGlnGluMetValArg. 1337
26862 GACCTGCGCGCTTCCGGGACCGGAGCGCGCTGGCGAGTGTGCCG 26911
1338 ArgSerArgGluArg.....SerProPheGlnAspPr 1348
26912 AACTGGTGGCGAGGCGCTGCACGCGCGCGCGCGCGCGTGGAGAC 26961
1348 oGluAsnValArgTyrArgLysSerValIleThrIleTyrLysGlnThrSera 1365
26962 GCGGCGACCTGGCGGAGCTG.....ACAGCTGAAGGCGCCACCTGGC 27005
1365 sParGValAspLys...ThrLysAspGluMetGlnHisGlnAlaLeuVa 1380
27006 GAAGGTGTCTCGCGCGCGCGCGCGCGCACTCGACATCCGTTCTACT 27055
1380 LGuGlyAsnLeuAlaThrGlu 1387
27056 CGAGGTCACCGGAGACTGAT 27077
```

seq\_name: /cgn2\_6/prodata/2/lna/5B\_comb.seq:us-08-804-198-1  
seq\_documentation\_block:

```
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kubstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rostock, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14045..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
; US-08-804-198-1

alignment_scores:
  Quality: 182.00      Length: 1649
  Ratio: 0.285        Gaps: 89
  Percent Similarity: 38.690  Percent Identity: 21.953

alignment_block:
US-09-736-968A-2 x US-08-804-198-1 ..
Align seg 1/1 to: US-08-804-198-1 from: 1 to: 44377
2 AaalaSerGluArgAlaPheAla.HisLysIleAsnArgThrVala 18
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23189 GCCTGCTCGACGACGATCAACGCGACCGCGCTCATCCGCGCTCTC 23238
18 laalagluvalarglysglnvalSerArgGluArgSerGlySerProHis 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
23239 TCGCT.....GCTGGACTGTGCACCGACGCGAGCGCGGACCGCCGC 23282
35 SerSerArgArgCysSerSerSerLeuGlyValProLeuThrGluVala 51
:::|||||
23283 CGTGGCCGTGCG..... 23294
51 lGluProLeuAspPheGluAspValLeuLeuSerArgProProAspAlaG 68
||| |||
23295 .....CGTACCGCCACCTCTCCCTC 23316
68 lndProGlyProLeuArgAspLeuValGluPheProAlaAspAspLeuGlu 84
:::||||| |||
23317 GTCCAGGCCCTGGCCGA..... 23333
85 LeuLeuLeuGlnProArgGluCysArgThrThrGluProGlyIleProIy 101
|||||:|||||:|||||:
23334 .....CCTGGCAGCACCGCACCGCT..... 23354
101 sAspGluLysLeuAspAlaGlnValArgAlaValaGluMetTyrIleG 118
:::|||||:|||||:|||||:
23355 .....GTGGACGTTCACCTGGCGCGGT.....CGCCACCGCC 23388
118 lAspThrValIleValHisArgArg..... 126
23389 CCGCAGCAACTGCCGTGCACCGCGGTGCCCACTGTGGGGCTGGGGCCG 23438
127 .....TyrGlnTyrIle 130
23439 GTGCGCGCGCTGAGCTCCGAGGTGTGGGGCGGCGCTTCACGACCTTC 23488
130 useAlaAlaTyrSerProValThrThrAspThrGlnArgGluGlnL 147
|||||:|||||:|||||:
23489 CCGCGGGCCCGACGCGCGGCTCTGGACCGTCCCGCGGT..... 23531
147 ysgLysLeuProArgGlnValPheGluGlnAspAlaSerGlyAspGluArg 163
|||||:|||||:|||||:
23532 .....CCTGCCGACAC.....CGGCGG 23549
164 SerGlyProGluAspSerAsnAspSerArgArg.....Glyse 176
|||||:|||||:|||||:
23550 CGAGGACGATCGCCGATCGATGGGAGGCGCTTCGGCGCGCGCGGTCC 23599
176 rGlySerProGluAspThrProArgSerSerGlyAlaSerSerIlePheA 193
:::||||| |||
23600 TCGGGAACCGCGCGGCTCCGCGCC..... 23627
193 spleuArgAsnLeuAlaAlaAspSerLeuLeuProSerLeuLeuGlnArg 209
:::|||||:|||||:|||||:
23628 .....GGCTGGCGCGCGCGGCGACCGCTCT.....CATCGC 23660
210 AlaAlaProGluAsp..... 214
|||||:|||||
23661 CGGCGACTCGACGAGGTGCCCGCGGACGTGTCCGCTCTCTCGAGG 23710
215 .....ValAspA 217
23711 ACGGCGCGGACCGCGTGTGCTGGCGGACCGACCGCCCGCGGCGGCG 23760
217 rGArgAsnGluThrLeuArgArgGlnHisArgProProAlaLeuLeuThr 233
|||||:|||||:|||||:|||||:
23761 GCGCGCCCGGACTGACCGGCGTCTCCCTGTCGCCGTGGCGTGCAGCT 23810
234 LeuTyrPro.....AlaProAspGluAspGluAlaValG 245
||| |||
23811 CACCGACCGCGCGGACGTGGCGCGCTGCTCGACGACGACCGCCACCG 23860
245 uArgCysSerArgProGluPro..... 252
:::|||||:|||||:
23861 TCGCGGTGACGCGCGCGCTGTGGTGGCGCGCGCTGCGCGGAGAG 23910
253 ..ProArgGluHisPheGlyGlnArgIleLeuValLysCysLeuSerIle 268
|||||:|||||:|||||:|||||:
23911 GCACCGCGGACAT...CGCGCGCGCGCTGCGCGCAAGACACGCGCGC 23957
```

```

268 uLysPheGluIleGluProIlePheGlyIleLeuAlaLeuTyrA 285
      :::::::::: ||| ||| ||| |||
23958 CGGCCACCTGTCGACCTGGCGCGCGCGCTCGACCGCGT... 24003
      :::::::::: ||| ||| ||| |||
285 sPValArgGluLysLysIleSerGluAspPheTyrPheAspLeuAsn 301
      :::::::::: ||| ||| ||| |||
24004 .....GTGCTGTTCTCTCGCTCGCTCGGAGTGTGG 24033
      :::::::::: ||| ||| ||| |||
302 SerAspSerMetLysGlyLeuLeuArgAlaHisGlyThrHis...ProAl 317
      :::::::::: ||| ||| ||| |||
24034 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCGACG 24083
      :::::::::: ||| ||| ||| |||
317 aIleSerThrLeuAlaArgSerAla.....IlePheSerValThrT 331
      :::::::::: ||| ||| ||| |||
24084 GCTGGCGACAGCGCGCGCGCGCGCGCGCGCTTCCTCGTGGCGCT 24133
      :::::::::: ||| ||| ||| |||
331 YrPro..... 332
      :::::::::: ||| ||| ||| |||
24134 GGAGCGCTGGCGCGGAGCGACCGCGCGCGCGCGCGAGCTTC 24183
      :::::::::: ||| ||| ||| |||
333 .....SerProAspIlePheLeuAlaI 340
      :::::::::: ||| ||| ||| |||
24184 CTCAGCGCGCGCGCGCTGCTCCCTCGACCCGAC...CAGCGGCTGG 24230
      :::::::::: ||| ||| ||| |||
340 eLysLeuGluLysValLeuGlnGlnGlyAspIleSerGluCys...Cys.. 355
      :::::::::: ||| ||| ||| |||
24231 GACCTGCGCGCATGCTGGAGCGCGG...AGCGCTGCGGTGG 24274
      :::::::::: ||| ||| ||| |||
356 .....GluProTyrMetValLeuLysGluValAspThrA 367
      :::::::::: ||| ||| ||| |||
24275 TCGCGGAGCTGAGTGGAGCGCGCTTCGCGCTCTACACCTGGTGGCT 24324
      :::::::::: ||| ||| ||| |||
367 IalysAsnLysGluLysLeuGluLysLeuArgLeuAlaIaGluGlnPhe 383
      :::::::::: ||| ||| ||| |||
24335 CCGCGCGCTCTTCGACGACATCCGCGAGCTGCGCGCTCGCGCGG 24374
      :::::::::: ||| ||| ||| |||
384 CysThrArgLeuGlyArgTyrArgMetProPheAlaThrPheAlaValH 400
      :::::::::: ||| ||| ||| |||
24375 CGAACTCGCGCGGACGACGAGACTCGAC.....CA 24406
      :::::::::: ||| ||| ||| |||
400 sLeuAlaAsnIleValSerSerAlaGlyGlnLeuAspArgAspSerAsps 417
      :::::::::: ||| ||| ||| |||
24407 CCTC.....CGAACTCGCGCGAGCTGACGCGCGAGTCCGCG 24444
      :::::::::: ||| ||| ||| |||
417 eArgLysGluArgArgProAlaThrPheAspArgArgGlyPro 433
      :::::::::: ||| ||| ||| |||
24445 CACAAGCGGACGCGACCTGCTCGCGCTGTGCGCGCACACGCGCGCG 24494
      :::::::::: ||| ||| ||| |||
434 GlnAspArgAla..SerSerGlyAspAspAlaCysSerPheSerGlyPheA 450
      :::::::::: ||| ||| ||| |||
24495 CGTCTCGGACAGTCTCGCGCGGAGGTGAGCAGCGCGCGCGCTTCC 24544
      :::::::::: ||| ||| ||| |||
450 rGrProAla.....ThrLeuThrValThrAsnPhelysGlnGlu 463
      :::::::::: ||| ||| ||| |||
24545 GCGACCTCGGCTTCGACTCGCTGACGCGCTCGAAGT... 24582
      :::::::::: ||| ||| ||| |||
464 ALGluArgLeuSerAspGluAspLeuPheLys.....PheLe 476
      :::::::::: ||| ||| ||| |||
24583 CGGACCGGCTCAGCACACCGCGCTCAAACTGCCACCTCCCTGTG 24632
      :::::::::: ||| ||| ||| |||
476 uAlaAspMetArgArgProSerSerLeuLeuArgArgLeu..... 489
      :::::::::: ||| ||| ||| |||
24633 CTTCGACCACTCCAGCGCGCGCTCGCGCGCGCTGAGGAAC 24682
      :::::::::: ||| ||| ||| |||
490 .....Arg..ProValThrAlaGlnLeu 497
      :::::::::: ||| ||| ||| |||
24683 TCCTCGCGCGGAACGACCGCGCGCGCGCGCGCGCGCGA 24732
      :::::::::: ||| ||| ||| |||
497 sIleAspIleSerProAlaProGluAsn.....ProH 508
      :::::::::: ||| ||| ||| |||
24733 CGGACGAGAGCGCATTCGCATTCATGGCATGGCTCGCGCTGCGCGG 24782
      :::::::::: ||| ||| ||| |||
508 tSPheCysLeuSerProGluLeuLeu.....HisIleLysPro 520
      :::::::::: ||| ||| ||| |||
24783 CGGGGTCACTCCCGAGGACTGTGGAGCTGTGACCGTGGAGCG 24832
      :::::::::: ||| ||| ||| |||
521 TyrProAspProArgGlyArgProThrLysGluIleLeuGluPheProAl 537
      :::::::::: ||| ||| ||| |||
24833 ACGCATACACCCCTTCGCCGACACCGGGGAGGAGCAACAGACCTTC 24882
      :::::::::: ||| ||| ||| |||
537 a.ArgGlu.....ValTyrAlaProHisThrSerTyrAlaAsn 549
      :::::::::: ||| ||| ||| |||
24883 TACGACCCCGACCCCGACTCGCGCGGACACCACTACCTGCGCGAGG 24932
      :::::::::: ||| ||| ||| |||
550 LeuLeuTyrValTyrProHisSerLeuAsnPheserArgGlnLys 566
      :::::::::: ||| ||| ||| |||
24933 .....CGGTTCTCGACAGACCGCGCG 24955
      :::::::::: ||| ||| ||| |||
566 rValArgAsnLeuAlaValArgValGln.....TyrMetThrGlyGlu 580
      :::::::::: ||| ||| ||| |||
24956 AGTTGACACCCCGCTTCCTCGCATCAGCCCGCGAGCGCGCTGCGCATG 25005
      :::::::::: ||| ||| ||| |||
581 AspProSerGlnAlaLeuProValIlePheGlyLysSerSerCysSerG 597
      :::::::::: ||| ||| ||| |||
25006 GACCCGACGAGCGCGCTGATCTG.....GACGCTCTGGGAGTC 25046
      :::::::::: ||| ||| ||| |||
597 uphethrArgGluAlaPheThrProValValTyrHisAsnLysSerProG 614
      :::::::::: ||| ||| ||| |||
25047 CTTCGAACGCGCGCGCATCGACCCGCTGCACTGCGCGCGCAC... 25089
      :::::::::: ||| ||| ||| |||
614 IuphetyrGluGluPheLysLeuHisLeuProAlaCysValThrGluAsn 630
      :::::::::: ||| ||| ||| |||
25089 ..... 25089
      :::::::::: ||| ||| ||| |||
631 HisHisLeuLeuPheThrPheThrHisValSerCysGlnProArgProG 647
      :::::::::: ||| ||| ||| |||
25090 .....CGACCGG 25097
      :::::::::: ||| ||| ||| |||
647 yThrAlaLeuGluThrProValGlyPheThrThrIleProLeuLeuGln 663
      :::::::::: ||| ||| ||| |||
25098 GGTCTTCGTGCGCACCAAC...GGACAGCATCATGCGCGCTCTCCAGG 25144
      :::::::::: ||| ||| ||| |||
664 .....HisGlyArgLeuArgThrGlyProPheCysLeu 674
      :::::::::: ||| ||| ||| |||
25145 ACGGCGACGAGACTTCGACGCGCTACATCGCCACGCG... 25182
      :::::::::: ||| ||| ||| |||
675 ProValSerValAspGlnProProSerTyrSerValLeuThrProAs 691
      :::::::::: ||| ||| ||| |||
25183 .....AACTCGCGACGCGAGTCCGCGCG 25208
      :::::::::: ||| ||| ||| |||
691 rValAlaLeuProGlyMetArgTyrValAspGlyHisLysGlyValPheS 708
      :::::::::: ||| ||| ||| |||
25209 GCTTCCTTAC.....GTCCTCG 25225
      :::::::::: ||| ||| ||| |||
708 eValGluLeuThrAlaValSerSerValHisProGlnAspProTyrLeu 724
      :::::::::: ||| ||| ||| |||
25226 GACTGGAGGCGCGCGCTCAC... 25248
      :::::::::: ||| ||| ||| |||
725 AspLysPhePheThrLeuValHisValLeuGluGluGlyAlaPheProPh 741
      :::::::::: ||| ||| ||| |||
25248 ..... 25248
      :::::::::: ||| ||| ||| |||
741 eArgLeuLysAspThrValLeuSerGluGlyAsnValGluGlnLeuA 758
      :::::::::: ||| ||| ||| |||
25249 .....GTGACACCGCGCTGCT 25264
      :::::::::: ||| ||| ||| |||
758 rGAlaSerLeuAlaLeuArgLeuAla..... 767
      :::::::::: ||| ||| ||| |||
25265 CGGCTCTCTGCGCGACTGCACCTGGCGGTGCACTACCTCGCGCGCG 25314
      :::::::::: ||| ||| ||| |||
768 .....SerProG 770
      :::::::::: ||| ||| ||| |||
25315 GAATGCGACTACGCGCTCGCGGCGGCGCACGCTGATGTCCACCCGGA 25364
      :::::::::: ||| ||| ||| |||
770 uProLeuValAlaPheSerHisHisValLeuAspLysLeuValArgLeu 787
      :::::::::: ||| ||| ||| |||

```

```
25365 GAGCTGTGTGGAGTTCGCCCTG.....CAGCGAG 25393
      ||||| |||:.....
787 aAlIeArPrOProIleIleSerGIyGlnIleValAsnLeuGIyArGIy 803
      ||| ||| |||:.....
25394 CGGTGTGCGCGGACAGCGCCGACAGCGGTGTGCGGAGCGCGCGACGGG 25443
      ||| ||| |||:.....
804 ALAPheGIuLAmAtAlAHISValValSerLeuValHISArGIySerLeuI 820
      ||| ||| |||:.....
25444 GTGCGTCTCCCGAGGAGCGGGGATGCTGTGTGAGCGGCTGTGCGA 25493
      ||| ||| |||:.....
820 uAlAlAGIAspHLArGIyHIScysPrOgInLeuAlAlATyValH 837
      ||| ||| |||:.....
25494 GCGCGACAG.....AAGGCGCAT...CGGTACTGGCGGTGTGTCGGG 25534
      ||| ||| |||:.....
837 tSTyRAlAPheArGIyLeuPrOgLy.....ThrGIuPrOser 848
      ||| ||| |||:.....
25535 GCAGTGCCTTCACACAGAGCGGTGCCAGCAGCGCTCACCGACCCACG 25584
      ||| ||| |||:.....
849 LeuPrO.....AspGIyAl 853
      ||| ||| |||:.....
25585 GGGCGCGCCAGAGCGGGTGTATACGGAGCGGCTGGCGAGCGGGGCT 25634
      ||| ||| |||:.....
853 aPrOPrOValThyAlGIAlAlAlAlArGIySerGIyArGIy 870
      ||| ||| |||:.....
25635 GACGCGCGCGACGTGAGCGGGTGCAG...GGCAGCGACACCGGACGC 25681
      ||| ||| |||:.....
870 rO..... 870
25682 CGCTCGGCGACCCATCGAGCGCGCGCTGCTGCCACGTACGCGCGG 25731
      ||| ||| |||:.....
871 .....AlAserLeuTyRLeuAlArGIySerGIySerLeuI 883
      ||| ||| |||:.....
25732 GACCGCGCGACGCGCGCTGTGTGTGCTGTGCTGTGAG..... 25770
      ||| ||| |||:.....
883 rSerAsPrOAspLeuAlValAlArPrOgIySerValAspAspGIuValS 900
      ||| ||| |||:.....
25771 .TGCATCATGCGGACACACCGCGCGCGCGGTGTGCGCGGGGTATGA 25819
      ||| ||| |||:.....
900 eArGIyIleuAlAserGIyLeuHISGIuGIuLeuAlAlGIuGIuTrp 916
      ||| ||| |||:.....
25820 AGAGTGTGTGCGG.....CTGGCGCACGCGCTGCGCGCACCGCTG 25863
      ||| ||| |||:.....
917 ValAlAserSerAlValArGIy.....GIuAl 926
      ||| ||| |||:.....
25864 CACCGGTGCAGCGCGCTGTCCAGGATGTGGAGCGCGCGCGCTGGA 25913
      ||| ||| |||:.....
926 aIleuGIuHISAlA.....TrpPhePheGIuLeuMetValIyS 941
      ||| ||| |||:.....
25914 GTTGTGTGAGAGCGCGCGCTGTGCTG..... 25941
      ||| ||| |||:.....
941 eRMeAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 957
      ||| ||| |||:.....
25942 .....CAGCGCGCGAGGGGCGCGCGCGG 25965
      ||| ||| |||:.....
958 LeuArGIyPhePrOgLyArGIyLeuAspAlIeThrAlAlAlAlAlAl 974
      ||| ||| |||:.....
25966 .....GCGGCGATCTCTCTC 25979
      ||| ||| |||:.....
974 rValGIyLeuGIuValIleThrArGIyAlHISIySAspValGIuLeuAlA 991
      ||| ||| |||:.....
25980 GTTGGCGATCTCAGCGG...ACCAACGCGCACCTGTCTCATCGAGACCG 26026
      ||| ||| |||:.....
991 LuHISLeuAsnAlAserLeuAlAPhePheLeuSerAspLeuLeuSerLeu 1007
      ||| ||| |||:.....
26027 CGGAGCCACCGCG.....CCGCAACTCTCTC..... 26052
      ||| ||| |||:.....
1008 ValAspArGIyGIyPheValPheSerLeuValArGIAlAlAlAlAlAl 1024
      ||| ||| |||:.....
26052 ..... 26052
1024 lAlAlArGIyLeuGIuInsSerSerPrOAsnPrOAlAlAlAlAlAlAl 1040
      ||| ||| |||:.....
```

```
26053 .....GCGCGCGAACCAGCGCGCGCGCGCGCGCT 26083
1041 .....ArgMeGIuPheThrArGIyLeuCysSerHISGIuHISTyRVal 1055
      ||| ||| |||:.....
26084 GGTCCGAGGAGGTGTGGACAGAGTGAACCGTGCCTGATGATGTCTCCG 26133
      ||| ||| |||:.....
1055 lThrLeuAsnLeuPrOgCys.....ProLeuSerPrOProA 1068
      ||| ||| |||:.....
26134 CACACAGACGCGCGCTGTGAGCAGCGCGCGCGCGCGCTGTGCGGACCT 26183
      ||| ||| |||:.....
1068 lAserPrOserPrOserValSerSerThrThrSerGIuSerSerThrPhe 1084
      ||| ||| |||:.....
26184 GCTGCGCCACCGCAGCTGACCGCGCGCGCGCGCGCTGTGCGA..... 26224
      ||| ||| |||:.....
1085 SerSerGIuAlAProAspPrOgLyValThrSerMetPheGIuLeuSerGI 1101
      ||| ||| |||:.....
26225 CCTCATACACCGCGCGCGCG..... 26247
      ||| ||| |||:.....
1101 yPrOAPheArGIyGlnHISPheLeuAlGIyLeuLeuThrGIuLeuAl 1118
      ||| ||| |||:.....
26248 ....TTCAGACAGCGGCGCGCGCTGTGCGCGAGAACTTCACGAGACTGA 26293
      ||| ||| |||:.....
1118 lAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 1134
      ||| ||| |||:.....
26294 TCGCGCGCTGTGAGCAGCTGTGTGAAGC..... 26322
      ||| ||| |||:.....
1135 AlAlIleSerAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 1150
      ||| ||| |||:.....
26323 .....CGACCGACCGCGCTGTGTGTGCGCGCGCGCGCGCGCTC 26363
      ||| ||| |||:.....
1151 ArGIyAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 1167
      ||| ||| |||:.....
26364 CGA.....CCAGGTGTGT...CGTCTTCCCGCGG 26392
      ||| ||| |||:.....
1167 oLeuLeuSerIleAlArGIyAspThrLeuPrOArGIyLeuHISAspPheAlA 1184
      ||| ||| |||:.....
26393 AGGCTGTGAGTGGCGCGAGATGCGCGAGG.....GCTG 26427
      ||| ||| |||:.....
1184 LuGIyPrOgIyGlnArGIySerArGIyAlAlAlAlAlAlAlAlAlAlAl 1200
      ||| ||| |||:.....
26428 CTGGCGCGCTGTGAGCGCTGTGCTC.....CTT 26456
      ||| ||| |||:.....
1201 GluGIyGlnGIyAspIleAlAlGIyThrIleAsnPrOserValAlAlAl 1217
      ||| ||| |||:.....
26457 CCGTGAAGACCGCGCGCGCTGTGAGCTGTGCGCGCGCGCGCGCGCT 26506
      ||| ||| |||:.....
1217 aIleAlAlGIyGIyPrOLeuAlAPrOgIySerArGIyAlAserIleSerGI 1234
      ||| ||| |||:.....
26507 GGTGCGTCTGTGAGTACTGTGCGCGGAGACCGCGCGCGCTGTGCGA 26555
      ||| ||| |||:.....
1234 LyPrOPrOThrAlAserArGIyCysAlAlAlAlAlAlAlAlAlAlAl 1250
      ||| ||| |||:.....
26556 .....CCGGTGTGACGTGTGTGACCGCGCTGTGTACCATGAT 26594
      ||| ||| |||:.....
1251 ArGIyThrLeuAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 1267
      ||| ||| |||:.....
26595 GGT.....CTGCTGTGCGAGAGCTGTGCTGTGCGGTGTGAGACCG 26638
      ||| ||| |||:.....
1267 lAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 1273
      ||| ||| |||:.....
26639 CCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 26688
      ||| ||| |||:.....
1274 ThrAspLeuThrLeuPrOgInLeuGIyArGIyLeuAspLeu..... 1287
      ||| ||| |||:.....
26689 GCGCGCTGTGAGCTGTGAGAGCGCGCGCGCGCGCGCTGTGCGAGCCA 26738
      ||| ||| |||:.....
1288 LeuTyRLeuCysLeuAlAlAlAlAlAlAlAlAlAlAlAlAlAlAlAl 1304
      ||| ||| |||:.....
26739 GCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 26773
      ||| ||| |||:.....
1304 LuArGIyLeuSerLeuThrPheIySISerLeuAspMetIySAlArGI 1320
      ||| ||| |||:.....
26774 TCGCGGTGTGAC.....CTGTGTGAGCGCGAGCTGTGTGTGTGTGT 26811
      ||| ||| |||:.....
```

```

1321 LeuGluGlu..... 1323
      |||||
26812 CTGGAGCCCTGGAGCGACCGGCTCGCCGCGCCGTCACGGCCCCGA 26861
1324 .....AlaIleLeuGlyThrIleGlyAlaArgGlnGluMetValArg. 1337
      |||||
26862 GACCTGGCGCGGTCTCCGGGGACCCGACGCCCTGGGGAGCTGTGTGCCG 26911
1338 ArgSerArgGluArg.....SerProPheGlyAsnPr 1348
      |||||
26912 AACTCGGTGGGAGGGGCTGCACGCCGCCCATCCCGCGGTGCACACC 26961
1348 OGluAsnValArgTPrArgLysSerValThrHisTrpLysGlnThrSerA 1365
      |||||
26962 GCCGGGCACTCGCCGAGTCG.....ACACGCTGAGAGCCCACTCTGCG 27005
1365 sPARgValAspLys...ThrLys.AsprGluMetGlnHisGluAlaLeuVa 1380
      |||||
27006 GAAGGTCTCTGCGCCGCTCGCGCCGCGCACTCCGACATCCGTTCTACT 27055
1380 lGluGlyAsnLeuAlaThrGlu 1387
      |||||
27056 CGACGTACACCGGAGGACTGAT 27077

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 02:45:43 ; Search time 899.83 Seconds  
(without alignments)  
12158.051 Million cell updates/sec

Title: US-09-736-968A-1  
Perfect score: 6372  
Sequence: 1 gagcgagagacatgctgc.....aagcaaaaaaaaaaaaaa 6372

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues  
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N.Geneseq\_032802:\*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6372	100.0	6372	22	AAH43851
2	2140	33.6	2148	22	AAH43850
3	2132.8	33.5	6828	22	AAH41934
4	1729.4	27.1	7215	22	AAH43851
5	1465.2	23.0	1652	22	AAH43851
6	1446.6	22.7	4143	22	AAH41911
7	1446.6	22.7	4144	22	AAH41912
8	1446.6	22.7	4200	22	AAH41913
9	1195.8	18.8	3173	22	AAH41952

10	1031.8	16.2	4027	22	AAH43851	Human DNA associat
11	867.4	13.6	2014	22	AAH43851	Human DNA associat
12	829.4	13.0	1273	21	AAH43851	Human secreted pro
13	468.6	7.4	586	21	AAH43851	Human OREF ORF118
14	463.4	7.3	981	23	AAH43851	DNA encoding novel
15	448	7.0	2427	22	AAH43851	Human polynucleoti
16	407.8	6.4	2433	22	AAH43851	Human polynucleoti
17	407.4	6.4	9389	23	AAH43851	Drosophila melanog
18	385.4	6.0	966	24	AAH43851	DNA encoding novel
19	356.6	5.6	1834	23	AAH43851	CDNA sequence #412
20	322.4	5.1	3899	22	AAH43851	Human CDNA encodin
21	284	4.5	300	21	AAH43851	Human colon cancer
22	266.8	4.2	5589	23	AAH43851	Drosophila melanog
23	255.2	4.0	6816	21	AAH43851	Human OREF ORF79 p
24	255.2	4.0	7506	22	AAH43851	Angiogenesis assoc
25	243.8	3.8	417	21	AAH43851	Human OREF ORF30 p
26	238.2	3.7	321	22	AAH43851	Human CLASP-7 intr
27	238.2	3.7	512	22	AAH43851	Human foetal liver
28	238.2	3.7	3023	22	AAH43851	Human CDNA sequenc
29	237.6	3.7	4420	22	AAH43851	Human polynucleoti
30	232	3.6	232	22	AAH43851	Human foetal liver
31	225.4	3.5	5688	21	AAH43851	Human CLASP relate
32	225.4	3.5	5688	21	AAH43851	DNA encoding a hum
33	225.4	3.5	7277	24	AAH43851	Human CDNA encodin
34	216.6	3.4	281	22	AAH43851	Human CLASP-7 intr
35	213.8	3.4	1605	22	AAH43851	Human CDNA associat
36	210	3.3	5214	21	AAH43851	Mouse CLASP-1 nucl
37	210	3.3	5214	21	AAH43851	DNA encoding a mur
38	199.8	3.1	2036	22	AAH43851	Human CDNA sequenc
39	194.8	3.1	2610	22	AAH43851	Human CDNA encodin
40	194.8	3.1	3472	22	AAH43851	Human CDNA encodin
41	191.6	3.0	4393	22	AAH43851	Human CDNA encodin
42	191.6	3.0	6454	22	AAH43851	Human polynucleoti
43	189.4	3.0	441	22	AAH43851	Human CLASP-7 intr
44	184.8	2.9	271	22	AAH43851	Human foetal liver
45	184.8	2.9	512	22	AAH43851	Human foetal liver

## ALIGNMENTS

RESULT 1	AAH43851	standard; CDNA: 6372 BP.
ID	AAH43851	
XX	AAH43851:	
AC	04-SEP-2001	(first entry)
XX		
DT	Human CLASP-7 encoding CDNA sequence SEQ ID NO:1.	
DE		
XX	Human; CLASP-7; cadherin-like asymmetry protein; immune response;	
KW	neuroprotective; antidiabetic; immunosuppressive; antihypertensive;	
KW	antiarthritic; hypotensive; anti-HIV; cytostatic; immunostimulant;	
KW	antianemic; antiinflammatory; ophthalmological; nephroprotective;	
KW	antihypertensive; antidiabetic; antiallergic; antibacterial; gene therapy;	
KW	chromosome 19q13.2; autoimmune disease; multiple sclerosis; toxemia;	
KW	juvenile diabetes; rheumatoid arthritis; puritic urticarial papule;	
KW	hypertension; Rh incompatibility; ss.	
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	13..6156
FT		/*tag= a
XX		/product= "CLASP-7"
PN	WO200142295-A2.	
XX		
PD	14-JUN-2001.	
XX		
PF	13-DEC-2000; 2000WO-US34152.	
XX		

PR 13-DEC-1999; 99US-0170453.  
PR 14-JAN-2000; 2000US-0176195.  
PR 14-FEB-2000; 2000US-0182296.  
PR 11-APR-2000; 2000US-0196267.  
PR 11-APR-2000; 2000US-0196460.  
PR 11-APR-2000; 2000US-0196527.  
PR 11-APR-2000; 2000US-0196528.  
PR 11-APR-2000; 2000US-0547276.  
PR 13-OCT-2000; 2000US-0240503.  
PR 13-OCT-2000; 2000US-0240508.  
PR 13-OCT-2000; 2000US-0240539.  
PR 13-OCT-2000; 2000US-0240543.  
XX

PA (ARBO-) ARBOR VITA CORP.

PI Lu P, Garman JD, Candia AF;

DR WPI; 2001-381641/40.

XX P-PSDB; AAB99541.

PT Novel cadherin-like asymmetry protein-7 and polynucleotides encoding  
PT the polypeptide, useful for treating autoimmune disease,  
PT hypersensitivity, preventing transplant rejection by modulating immune  
PT response

XX Claim 1; Fig 5; 151pp; English.

XX The present invention describes a human cadherin-like asymmetry protein  
CC (CLASP), designated CLASP-7. The CLASP-7 protein (I) and its encoding  
CC nucleotide sequence (II) have activities including: neuroprotective;  
CC anti-diabetic; immunosuppressive; antirheumatic; antiallergic; anti-HIV;  
CC hypotensive; cytostatic; immunostimulant; antineoplastic; anti-inflamatory;  
CC ophthalmological; nephrotoxic; antihypertensive; antidiabetic; anti-inflammatory;  
CC antiallergic; and antibacterial. (II) and CLASP-7 antibodies (III) are  
CC useful for detecting the CLASP-7 polypeptide. (II) is useful for  
CC producing (I) by recombinant methods. (I) or its fragment are useful for  
CC inhibiting an immune response in a cell such as T cell or B cell. A  
CC pharmaceutical composition (C), comprising (I) or (II), can be useful  
CC for treating CLASP-7-mediated disease such as an autoimmune disease  
CC caused or exacerbated by increased activity of T helper cells. Autoimmune  
CC diseases which can be treated using (C) include multiple sclerosis,  
CC juvenile diabetes and rheumatoid arthritis. (I) is useful for treating  
CC toxemia or pregnancy induced hypertension, pruritic urticarial papules  
CC and Rh incompatibility. (I) is also useful as a diagnostic reagent for  
CC immune and other disorders, since diseases characterized by  
CC overproduction or depletion of lymphocytes in blood or other organs may  
CC be detected by monitoring the level of (I) or its mRNA. CLASP-7 has been  
CC mapped to the chromosomal location 19q13.2. The present sequence  
CC encodes the human CLASP-7 protein as given in the present invention.  
XX

XX Sequence 6372 BP; 1273 A; 2050 C; 1849 G; 1200 T; 0 other;

Query Match 100.0%; Score 6372; DB 22; Length 6372;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 6372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacgcgaagacatgctgctccctccgagcgccgcttcgcgcacacagatcaacagcagc 60  
DB 1 gacgcgaagacatgctgctccctccgagcgccgcttcgcgcacacagatcaacagcagc 60  
QY 61 gtcgcgcgacgaggtgcggaagcaggtgtcccggaacgacagtggtgtcccccacacgc 120  
DB 61 gtcgcgcgacgaggtgcggaagcaggtgtcccggaacgacagtggtgtcccccacacgc 120  
QY 121 aggcgcgtgacgacgtccctcgtgggttccacactgactgaagtgtgcagcccttgactt 180  
DB 121 aggcgcgtgacgacgtccctcgtgggttccacactgactgaagtgtgcagcccttgactt 180  
QY 181 gagagtgtaactctctgacgcgcgcacacagatctgagcccgggccctcctaaggaactgtga 240  
DB 181 gagagtgtaactctctgacgcgcgcacacagatctgagcccgggccctcctaaggaactgtga 240

QY 241 gattcccaagctgacttgagctgtctgcagcccggaatcgccgaccagag 300  
DB 241 gaattcccaagctgacttgagctgtctgcagcccggaatcgccgaccagag 300  
QY 301 cccggatccccaagatgtaaaactgtgacccaggtgagggcggtgagatgtat 360  
DB 301 cccggatccccaagatgtaaaactgtgacccaggtgagggcggtgagatgtat 360  
QY 361 attgagactgggtatcttccacagaaagtatacagtaactgtgacataagcccc 420  
DB 361 attgagactgggtatcttccacagaaagtatacagtaactgtgacataagcccc 420  
QY 421 gtcaacacagacacacagcggagcgacagaaaggcctcccccgcacagttcttgacag 480  
DB 421 gtcaacacagacacacagcggagcgacagaaaggcctcccccgcacagttcttgacag 480  
QY 481 gatgtctcttgagacgagaggtccggccctgaagacttcgaatcccggtgtctcg 540  
DB 481 gatgtctcttgagacgagaggtccggccctgaagacttcgaatcccggtgtctcg 540  
QY 541 ggtcccccgaagacacccctcggaagcagtggtcctctagcatcttcgacctgaggaac 600  
DB 541 ggtcccccgaagacacccctcggaagcagtggtcctctagcatcttcgacctgaggaac 600  
QY 601 ctgacgactgacatctgctcctctctgtagagcgggcgccccaagatgtgac 660  
DB 601 ctgacgactgacatctgctcctctctgtagagcgggcgccccaagatgtgac 660  
QY 661 cggcgcaatgaaaccttgagcggcagacacggcccgccctgcacaccttaccg 720  
DB 661 cggcgcaatgaaaccttgagcggcagacacggcccgccctgcacaccttaccg 720  
QY 721 gcaactgacgaggtgaaagccgtgaaacgctgtgacggccgaagacaccccgagac 780  
DB 721 gcaactgacgaggtgaaagccgtgaaacgctgtgacggccgaagacaccccgagac 780  
QY 781 ttgtgacaaagatctgtgtcaagtgtctgtcgtcaagtgtgagatgtaattgaccc 840  
DB 781 ttgtgacaaagatctgtgtcaagtgtctgtcgtcaagtgtgagatgtaattgaccc 840  
QY 841 atctctggacatctggtcctgtatgatgtgcggaagaagaagatctcgagaacttc 900  
DB 841 atctctggacatctggtcctgtatgatgtgcggaagaagaagatctcgagaacttc 900  
QY 901 tacttcgaactgaaactcgactccatgaagaagggtgtcttggtgctaatgtgacacccct 960  
DB 901 tacttcgaactgaaactcgactccatgaagaagggtgtcttggtgctaatgtgacacccct 960  
QY 961 gcaatctccacccctggccgctctgtccatctctctgtgacctaccctcaactgacatc 1020  
DB 961 gcaatctccacccctggccgctctgtccatctctctgtgacctaccctcaactgacatc 1020  
QY 1021 ttctctgtcatcaagtttgaggaaggtgtcttcacgaagaaggacatcagtgatgtctgag 1080  
DB 1021 ttctctgtcatcaagtttgaggaaggtgtcttcacgaagaaggacatcagtgatgtctgag 1080  
QY 1081 ccttacatgtgtgtgaaagaagtgtgacacagccaagaagaagaagaagttagaagaagctg 1140  
DB 1081 ccttacatgtgtgtgaaagaagtgtgacacagccaagaagaagaagaagttagaagaagctg 1140  
QY 1141 cgctctggcgcgacgagatctgtcacccgccttggtgcctacacgcatgtcccttcgcctg 1200  
DB 1141 cgctctggcgcgacgagatctgtcacccgccttggtgcctacacgcatgtcccttcgcctg 1200  
QY 1201 acgacgcgtgacttgccacacatcgttgagagagcgctgggacgttgagaccggagacttgac 1260  
DB 1201 acgacgcgtgacttgccacacatcgttgagagagcgctgggacgttgagaccggagacttgac 1260  
QY 1261 tcggaagggcagagcgccgacagccttgaaacagccgcgcttcgggggccccagagacgg 1320  
DB 1261 tcggaagggcagagcgccgacagccttgaaacagccgcgcttcgggggccccagagacgg 1320  
QY 1321 gcgagtagtgaggagacagcgcctgcagctctctgtgcttcgcgtccagccagctaaactgtc 1380



Db 1321 gcsagatagvggagcagccgctcgcagctctctctgctccgaccagccagctaacgtctc 1380  
QY 1381 acaaaactctttaagcagagagctgagcgaactgaacagagagacctcttaagtctctg 1440  
Db 1381 acaaaactctttaagcagagagctgagcgaactgaacagagagacctcttaagtctctg 1440  
QY 1441 gctgacatgagagccgcctcgtctccctgctgctgagcgaactaagctctctgtgactgcccagctc 1500  
Db 1441 gctgacatgagagccgcctcgtctccctgctgctgagcgaactaagctctctgtgactgcccagctc 1500  
QY 1501 aagatcgaactctctccgctcctctgaaaatccccaactctgctctccctcgagagtctt 1560  
Db 1501 aagatcgaactctctccgctcctctgaaaatccccaactctgctctccctcgagagtctt 1560  
QY 1561 catatcaagcccttaacccagagccccaagggccggccagcccaagagaattctgagatctccc 1620  
Db 1561 catatcaagcccttaacccagagccccaagggccggccagcccaagagaattctgagatctccc 1620  
QY 1621 gcccgagagctctatgccccccaataccagctaacaggaacctgtgtacgtgtacccggcac 1680  
Db 1621 gcccgagagctctatgccccccaataccagctaacaggaacctgtgtacgtgtacccggcac 1680  
QY 1681 agcctcaactttagcagacgcgcgcagaggtctccgctgscgaactctgctgagtgagtgatc 1740  
Db 1681 agcctcaactttagcagacgcgcgcagaggtctccgctgscgaactctgctgagtgagtgatc 1740  
QY 1741 atgacagagcagagagacccccaagcagagctctgcccagctaatcttgcgaagctcagcgtcag 1800  
Db 1741 atgacagagcagagagacccccaagcagagctctgcccagctaatcttgcgaagctcagcgtcag 1800  
QY 1801 gaatttaaccgcgagagccttcaacaccggtgtgtcttaacataaagaatcccccagatctaac 1860  
Db 1801 gaatttaaccgcgagagccttcaacaccggtgtgtcttaacataaagaatcccccagatctaac 1860  
QY 1861 gaggagatcaagctcgaatcttccagcctgtgctgacagagaacatcacactgtgtctaac 1920  
Db 1861 gaggagatcaagctcgaatcttccagcctgtgctgacagagaacatcacactgtgtctaac 1920  
QY 1921 ttctaacatgtcaagctgcccagcccgccggccggaactgtccctgagagacaacccgttgagctt 1980  
Db 1921 ttctaacatgtcaagctgcccagcccgccggccggaactgtccctgagagacaacccgttgagctt 1980  
QY 1981 acttgatctccaactcgtctgtaagacagcgccgctgaagagacccgctctgtctcccaatg 2040  
Db 1981 acttgatctccaactcgtctgtaagacagcgccgctgaagagacccgctctgtctcccaatg 2040  
QY 2041 tctgtggaacagccgcgcgcagctatctccgtgctcaaccccgatgtgagcttccgggc 2100  
Db 2041 tctgtggaacagccgcgcgcagctatctccgtgctcaaccccgatgtgagcttccgggc 2100  
QY 2101 atgagcctgggtgagagcgttacaagaagcggtgtacagtgtgagagctacaagccgctgtccct 2160  
Db 2101 atgagcctgggtgagagcgttacaagaagcggtgtacagtgtgagagctacaagccgctgtccct 2160  
QY 2161 gtagaacccccagagaccctactctgtaacaaattcttaaccctgtgtacagcttccgtgagag 2220  
Db 2161 gtagaacccccagagaccctactctgtaacaaattcttaaccctgtgtacagcttccgtgagag 2220  
QY 2221 gtagaccttccactcccggttcaaggaactgtgctgagcgaaggaaggtgagagagag 2280  
Db 2221 gtagaccttccactcccggttcaaggaactgtgctgagcgaaggaaggtgagagagag 2280  
QY 2281 ctgagcgagcagctctgacagacatgctgctgagcgaagcccgaaacccctgtgtccctcc 2340  
Db 2281 ctgagcgagcagctctgacagacatgctgctgagcgaagcccgaaacccctgtgtccctcc 2340  
QY 2341 caacacgtgtctgtaacaaagctcgtgctgtgtcatcaagggcccgatcaagtgagcag 2400  
Db 2341 caacacgtgtctgtaacaaagctcgtgctgtgtcatcaagggcccgatcaagtgagcag 2400  
QY 2401 attgtgaaccttgagccgtgtgagccttgaaggaatgagcccatgtgtagccttctaac 2460

Db 2401 attgtgaaccttgagccgtgtgagccttgaaggaatgagcccatgtgtagccttctaac 2460  
QY 2461 cggagaccttgagaggaagcccaagatgcccgcgtgcaactgcccacaagctggtctactagtc 2520  
Db 2461 cggagaccttgagaggaagcccaagatgcccgcgtgcaactgcccacaagctggtctactagtc 2520  
QY 2521 cactaagccttctgccttctctgagcactggaacccaagctcccggaatggggccctccagatg 2580  
Db 2521 cactaagccttctgccttctctgagcactggaacccaagctcccggaatggggccctccagatg 2580  
QY 2581 acaatgagagctgccaacacgagcccggtgtctgtgtccgcccgaagcctctacccgtggc 2640  
Db 2581 acaatgagagctgccaacacgagcccggtgtctgtgtccgcccgaagcctctacccgtggc 2640  
QY 2641 cgttccaaagacatcaagcaagcaacaaactgaaactcgcgtgagccctggtctgtgtgat 2700  
Db 2641 cgttccaaagacatcaagcaagcaacaaactgaaactcgcgtgagccctggtctgtgtgat 2700  
QY 2701 gacgaagcttcccgcaactcctgagcagaaagctgtcttaacgaagagctgtgcttgcagatg 2760  
Db 2701 gacgaagcttcccgcaactcctgagcagaaagctgtcttaacgaagagctgtgcttgcagatg 2760  
QY 2761 gtgttcaagcagcagtgccgctgacagagagcgaatcctccagagacgcctgtgtctctccag 2820  
Db 2761 gtgttcaagcagcagtgccgctgacagagagcgaatcctccagagacgcctgtgtctctccag 2820  
QY 2821 ctcaatgtgaagatlatagcgctgtgcaacctgtgcttgcagcagagctatgacacaccgcg 2880  
Db 2821 ctcaatgtgtgaagatlatagcgcgctgtgcaacctgtgcttgcagcagagctatgacacaccgcg 2880  
QY 2881 aagctgtgccttcccggaagccttccctgagagacatacctctgtgtgggtctctgtgggc 2940  
Db 2881 aagctgtgccttcccggaagccttccctgagagacatacctctgtgtgggtctctgtgggc 2940  
QY 2941 ctggaagctcaatcaacccgtgtcccaacaagaagtgtgagcgtgscgagacacccaacgcgcag 3000  
Db 2941 ctggaagctcaatcaacccgtgtcccaacaagaagtgtgagcgtgscgagacacccaacgcgcag 3000  
QY 3001 ctggcttctctccctcagtgaaactctctgtccctgtgtgagaccgggcttctgtcttcaacgtg 3060  
Db 3001 ctggcttctctccctcagtgaaactctctgtccctgtgtgagaccgggcttctgtcttcaacgtg 3060  
QY 3061 gtcggggcccaactaagaagcaggtgtgcccagcggctccaggtccgtcccttaatccagaagcc 3120  
Db 3061 gtcggggcccaactaagaagcaggtgtgcccagcggctccaggtccgtcccttaatccagaagcc 3120  
QY 3121 ctgcttgaacccctgacatgagaaatcaaccgcgaatcctgtgtgacggccaagacatacgtgac 3180  
Db 3121 ctgcttgaacccctgacatgagaaatcaaccgcgaatcctgtgtgacggccaagacatacgtgac 3180  
QY 3181 ctcaaacctccctgtgtgcccctgtgtcaactccagcctgcgcctccctctgtgtctcc 3240  
Db 3181 ctcaaacctccctgtgtgcccctgtgtcaactccagcctgcgcctccctctgtgtctcc 3240  
QY 3241 accaactcccaagaactccaacttctccagccaagcccggaacccaaggtgagaccaatg 3300  
Db 3241 accaactcccaagaactccaacttctccagccaagcccggaacccaaggtgagaccaatg 3300  
QY 3301 ttccgaactgagtgaaacatctcccgagcagaacacttccctaagctgtggctcctcgtgagaggag 3360  
Db 3301 ttccgaactgagtgaaacatctcccgagcagaacacttccctaagctgtggctcctcgtgagaggag 3360  
QY 3361 ctggaacttgagccctcgaacactgagagctgaaaggggcatcttctgtgtgcaaaagagccatc 3420  
Db 3361 ctggaacttgagccctcgaacactgagagctgaaaggggcatcttctgtgtgcaaaagagccatc 3420  
QY 3421 agtgcgtgtgcaagcctgtcatgtgtgccaatgaaactgaaaccccgctacgcgcggagccact 3480  
Db 3421 agtgcgtgtgcaagcctgtcatgtgtgccaatgaaactgaaaccccgctacgcgcggagccact 3480  
QY 3481 gtagaagctgtgtgagcggagcgtgtacctgcaactgtcttcgaattgcaagggatattctg 3540  
Db 3481 gtagaagctgtgtgagcggagcgtgtacctgcaactgtcttcgaattgcaagggatattctg 3540

QY 3541 ccaagctgcatgacttctgtgagggccaggtcaagcgttcaagacttgccataatgctt 3600  
|||||  
Db 3541 ccaaggtgcatgacttctgtgagggccaggtcagcgtgtcaagacttgccataatgctt 3600  
QY 3601 gactcagaacacagaagcgagagggacatttgcggttaacataccctctgtgccaatg 3660  
|||||  
Db 3601 gactcagaacacagaagcgagagggacatttgcggttaacataccctctgtgccaatg 3660  
QY 3661 gccattctgtgccccccagcccttgctcccgggccaagcattccccaaggggccaaca 3720  
|||||  
Db 3661 gccattctgtgccccccagcccttgctcccgggccaagcattccccaaggggccaaca 3720  
QY 3721 acgggtctctcgccagagctgtgtgcctctctgtctgtgagtcaagccggaacctgtgcgctgt 3780  
|||||  
Db 3721 acgggtctctcgccagagctgtgtgcctctctgtctgtgagtcaagccggaacctgtgcgctgt 3780  
QY 3781 gtgtctgtgtgtgtctgtgaaaaaacacccgagccggcgctctctcagacgctgggccaactgtac 3840  
Db 3781 gtgtctgtgtgtgtctgtgaaaaaacacccgagccggcgctctctcagacgctgggccaactgtac 3840  
QY 3841 acacatccccagctgtggaacgtctgtctgtggaattgttaccttggctgtgctgtgtgag 3900  
|||||  
Db 3841 acacatccccagctgtggaacgtctgtctgtggaattgttaccttggctgtgctgtgtgag 3900  
QY 3901 tacaaagggaaaaaaggcctttgaaacgatalcaacagcctcaacattcaaaaaatctctgata 3960  
|||||  
Db 3901 tacaaagggaaaaaaggcctttgaaacgatalcaacagcctcaacattcaaaaaatctctgata 3960  
QY 3961 atgaaagcgcggtgtagaaggaagacattctgtgtgtacacatcggtgagctgtcagaagaatgctt 4020  
|||||  
Db 3961 atgaaagcgcggtgtagaaggaagacattctgtgtgtacacatcggtgagctgtcagaagaatgctt 4020  
QY 4021 cggcgaaagtctgtagaaggaagcccgcttctgtggaatccagaagaacgtgtgcgcggcgaaagac 4080  
|||||  
Db 4021 cggcgaaagtctgtagaaggaagcccgcttctgtggaatccagaagaacgtgtgcgcggcgaaagac 4080  
QY 4081 gtcaacacactgtgaagcaaacctcagaacccgtgtgagacagaacaaagataaattgtgaacac 4140  
Db 4081 gtcaacacactgtgaagcaaacctcagaacccgtgtgagacagaacaaagataaattgtgaacac 4140  
QY 4141 gaaggccttggtagaaggggaacccgtgcgaacccagagcaagccttggttctgtgacacacatg 4200  
|||||  
Db 4141 gaaggccttggtagaaggggaacccgtgcgaacccagagcaagccttggttctgtgacacacatg 4200  
QY 4201 gaagatacgtgtgacagacggtgtatgtcttcaagaagcccggaagacgcttctgtggggcagtg 4260  
|||||  
Db 4201 gaagatacgtgtgacagacggtgtatgtcttcaagaagcccggaagacgcttctgtggggcagtg 4260  
QY 4261 ctgaaaggttgtgtctgtacacagccttgggcagtgcccaagagtgcctctctcttgacagcatg 4320  
|||||  
Db 4261 ctgaaaggttgtgtctgtacacagccttgggcagtgcccaagagtgcctctctcttgacagcatg 4320  
QY 4321 ctggcgcccccaaggggcccttgtgtccaaagttcccggaactgtgttctgtagaggaacag 4380  
|||||  
Db 4321 ctggcgcccccaaggggcccttgtgtccaaagttcccggaactgtgttctgtagaggaacag 4380  
QY 4381 gaagctgtgtccgaacctgtgcctgtagcctcctaagacatgtgtgcagccgcatcagacac 4440  
|||||  
Db 4381 gaagctgtgtccgaacctgtgcctgtagcctcctaagacatgtgtgcagccgcatcagacac 4440  
QY 4441 atccgcacgacagccagcgcctctgttacctgtctcatgtgcagacaagaacttcgaatctggc 4500  
|||||  
Db 4441 atccgcacgacagccagcgcctctgttacctgtctcatgtgcagacaagaacttcgaatctggc 4500  
QY 4501 cacaacttggcccggtgtgtagagatgcaggttcacacatgtctctctgtccctgtgtggagag 4560  
|||||  
Db 4501 cacaacttggcccggtgtgtagagatgcaggttcacacatgtctctctgtgtggagag 4560  
QY 4561 acccagaacttcagtgaaagacacctgtgcagcttcaactaaaaacatccctacatgct 4620  
|||||  
Db 4561 acccagaacttcagtgaaagacacctgtgcagcttcaactaaaaacatccctacatgct 4620

QY 4621 gaagaggaacatgtgggtctgcgggacagacaccttcgcagagcaggtccaggaactgatgttc 4680  
|||||  
Db 4621 gaagaggaacatgtgggtctgcgggacagacaccttcgcagagcaggtccaggaactgatgttc 4680  
QY 4681 aaactgtacatgtatctctgaacgggacacgggtgaagatgaaggaacacccaggaagacctgtag 4740  
|||||  
Db 4681 aaactgtacatgtatctctgaacgggacacgggtgaagatgaaggaacacccaggaagacctgtag 4740  
QY 4741 atgtctcatcgacctcatgttacaagaattgccccgggggtctacccaggggtctaacccgaccttc 4800  
|||||  
Db 4741 atgtctcatcgacctcatgttacaagaattgccccgggggtctacccaggggtctaacccgaccttc 4800  
QY 4801 ctgacacctgtgtgcagaacatgtgcgggagacacgcgcgtgtggccaacacacgcgcgaagcc 4860  
|||||  
Db 4801 ctgacacctgtgtgtgcagaacatgtgcgggagacacgcgcgtgtggccaacacacgcgcgaagcc 4860  
QY 4861 gcccaagtgtcatgtgtgcaacgcggcccgccctctgtgtgagtgatccctcgccctgtctgaagac 4920  
Db 4861 gcccaagtgtcatgtgtgcaacgcggcccgccctctgtgtgagtgatccctcgccctgtctgaagac 4920  
QY 4921 cagcgccacacctgtgcgggtgtgtgcgttctctctcagaagaacatctacatccaaagtgctagag 4980  
|||||  
Db 4921 cagcgccacacctgtgcgggtgtgtgcgttctctctcagaagaacatctacatccaaagtgctagag 4980  
QY 4981 gagtccgcacatctccgaagacatccctgtgcggccagacgagagggctctctgcctccgggaag 5040  
Db 4981 gagtccgcacatctccgaagacatccctgtgcggccagacgagagggctctctgcctccgggaag 5040  
QY 5041 caattcacaatgtgaggtctgt 5100  
|||||  
Db 5041 caattcacaatgtgaggtctgt 5100  
QY 5101 ggcggtcctcagagagcggtgaatgaaggtctcaacaagaacccatcccatccctctgaaagcc 5160  
|||||  
Db 5101 ggcggtcctcagagagcggtgaatgaaggtctcaacaagaacccatcccatccctctgaaagcc 5160  
QY 5161 caacgtgtactcaagaagaagctgtgcgcggtgtgcagcgcaaacctgtcagagagccttaccacaa 5220  
Db 5161 caacgtgtactcaagaagaagctgtgcgcggtgtgcagcgcaaacctgtcagagagccttaccacaa 5220  
QY 5221 atcatgtcaccaagagttccgggtctgggagcggtgtgttcgggagagttatcccggtgtgtgtc 5280  
|||||  
Db 5221 atcatgtcaccaagagttccgggtctgggagcggtgtgtgttcgggagagttatcccggtgtgtgtc 5280  
QY 5281 taaggcgccacactccgggtgtgacctgtgagatgtgaagcagaggttgtgtacaaggggccaatcagc 5340  
|||||  
Db 5281 taaggcgccacactccgggtgtgacctgtgagatgtgaagcagaggttgtgtacaaggggccaatcagc 5340  
QY 5341 acgaagctgtgcagagatctcaacacggcctggaagagttctacacggagagattgtgcgac 5400  
|||||  
Db 5341 acgaagctgtgcagagatctcaacacggcctggaagagttctacacggagagattgtgtgcgac 5400  
QY 5401 gaagtcgtgtgagattatcaaaagactcttaacctgtggaacaagttccaagctgtgactcaaa 5460  
|||||  
Db 5401 gaagtcgtgtgagattatcaaaagactcttaacctgtggaacaagttccaagctgtgactcaaa 5460  
QY 5461 aaagcctacatccagatatacgtatgtgtgaacgtaactgtatacctcagcagctcaagac 5520  
|||||  
Db 5461 aaagcctacatccagatatacgtatgtgtgaacgtaactgtatacctcagcagctcaagac 5520  
QY 5521 cgggtgtgacctacttctgacggaactatgaggtcttcgcacatctcgttctgtcacgcgcttc 5580  
|||||  
Db 5521 cgggtgtgacctacttctgacggaactatgaggtcttcgcacatctcgttctgtcacgcgcttc 5580  
QY 5581 acgcgcgatatgtggcgacacacggggagctgtcccggaacacacaaagcgttaagaagcgtgtc 5640  
|||||  
Db 5581 acgcgcgatatgtggcgacacacggggagctgtcccggaacacacaaagcgttaagaagcgtgtc 5640  
QY 5641 agcacacgaacacgcttcccccatacatagaagctgtgacatccgtgtgtgtgtgtgtgtgtgtgtgt 5700  
|||||  
Db 5641 agcacacgaacacgcttcccccatacatagaagctgtgacatccgtgtgtgtgtgtgtgtgtgtgtgt 5700  
QY 5701 acggtgtgtgacgcaggtgtgaggtgtgcatcgtagagacatgtcagaagaagacacggtgagctg 5760

```

|||||
Db 5701 acggtcgtcagccagagtgaggtggtccatcgagagacatgcagaagaagacacggagctctg 5760
5761 gctcttcgacccagcagagcagccacccagatgctcaagatgctacagatgtgtctcaggcg 5820
5762 gctcttcgacccagcagagcagccacccagatgctcaagatgctacagatgtgtctcaggcg 5820
5821 tctgttagggccacacgtgaaacagaggtccctgtgaggtgtgcccaggtgttttagcagag 5880
5822 tctgttagggccacacgtgaaacagaggtccctgtgaggtgtgcccaggtgttttagcagag 5880
5881 atcccggaagaccccaagccttcctcgcatcacacaattgtggctcgtctcgaagagac 5940
5882 atcccggaagaccccaagccttcctcgcatcacacaattgtggctcgtctcgaagagac 5940
5941 tctcgaagaagatgtgagagatgcctcgtcggaagaaataagagcccttgatttgagccgag 6000
5942 tctcgaagaagatgtgagagatgcctcgtcggaagaaataagagcccttgatttgagccgag 6000
6001 aaggaagtaacacacgtgtgagcgtgagacgaactactgcgcgtcgtggagagctctgacccc 6060
6002 aaggaagtaacacacgtgtgagcgtgagacgaactactgcgcgtcgtggagagctctgacccc 6060
6061 ctgcttacccagcgtcgtcccgacgtgagcagccacccaccccgctcgaagactcc 6120
6062 ctgcttacccagcgtcgtcccgacgtgagcagccacccaccccgctcgaagactcc 6120
6121 ttgaacagagcaagtttcgaaagagcagaccttgagcccaagaagacaaagctgtacc 6180
6122 ttgaacagagcaagtttcgaaagagcagaccttgagcccaagaagacaaagctgtacc 6180
6181 tagaggaacacagcagcccgccctcagctgtctgtctgagagggaggtctgcctctgtgc 6240
6182 tagaggaacacagcagcccgccctcagctgtctgtctgagagggaggtctgcctctgtgc 6240
6241 ccactgggctgtgtgggtgacacacatgtaactgtgggctgtggccctcgtctgcc 6300
6242 ccactgggctgtgtgggtgacacacatgtaactgtgggctgtggccctcgtctgcc 6300
6301 catctgtgtgacatgactgtctcctctttaaattaaatggttttataagcaaaa 6360
6302 catctgtgtgacatgactgtctcctctttaaattaaatggttttataagcaaaa 6360
6361 aaaaaaaaaa 6372
6362 aaaaaaaaaa 6372

```

RESULT 2  
AAH43850  
ID AAH43850 standard; cDNA; 2148 BP.  
XX AAH43850;

04-SEP-2001 (First entry)

Preliminary human CLASP-7 encoding cDNA sequence.

Human; CLASP-7; cadherin-like asymmetry protein; immune response; neuroprotective; antidiabetic; immunosuppressive; antirheumatic; antitubercular; hypotensive; anti-HIV; cytosolic; immunostimulant; antianemic; antiinflammatory; ophthalmological; nephrotoxic; antihypertensive; antisthmatic; antiallergic; antibacterial; gene therapy; chromosome 19q13.2; autoimmune disease; multiple sclerosis; toxemia; juvenile diabetes; rheumatoid arthritis; pruritic urticarial papule; hypertension; Rh incompatibility; ss.

OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT 1..1932  
FT CDS /tag= a  
FT /partial

```

FT /product= "preliminary CLASP-7 protein"
FT /note= "no start codon given"
PD WO200142295-A2.
PD 14-JUN-2001.
PF 13-DEC-2000; 2000WO-US34152.
PR 13-DEC-1999; 99US-0170453.
PR 14-JAN-2000; 2000US-0176195.
PR 14-FEB-2000; 2000US-0182296.
PR 11-APR-2000; 2000US-0196267.
PR 11-APR-2000; 2000US-0196460.
PR 11-APR-2000; 2000US-0196527.
PR 11-APR-2000; 2000US-0196528.
PR 11-APR-2000; 2000US-0547276.
PR 13-OCT-2000; 2000US-0240503.
PR 13-OCT-2000; 2000US-0240508.
PR 13-OCT-2000; 2000US-0240539.
PR 13-OCT-2000; 2000US-0240543.
PA (ARBO-) ARBOR VITA CORP.
XX
XX Lu P, Garman JD, Candia AF;
XX WPI; 2001-381641/40.
XX P-PSDB; AAB99540.
XX
XX Novel cadherin-like asymmetry protein-7 and polynucleotides encoding
XX the polypeptide, useful for treating autoimmune disease,
XX hypersensitivity, preventing transplant rejection by modulating immune
XX response.
XX
XX Example 3; Fig 1; 151pp; English.
XX
XX The present invention describes a human cadherin-like asymmetry protein
XX (CLASP), designated CLASP-7. The CLASP-7 protein (I) and its encoding
XX nucleotide sequence (II) have activities including: neuroprotective;
XX antidiabetic; immunosuppressive; antirheumatic; antitubercular;
XX hypotensive; cytosolic; immunostimulant; antianemic; antiinflammatory;
XX ophthalmological; nephrotoxic; antihypertensive; antisthmatic;
XX antiallergic; and antibacterial. (II) and CLASP-7 antibodies (III) are
XX useful for detecting the CLASP-7 polypeptide. (III) is useful for
XX producing (I) by recombinant methods. (I) or its fragment are useful for
XX inhibiting an immune response in a cell such as T cell or B cell. A
XX pharmaceutical composition (C), comprising (I) or (II), can be useful
XX for treating CLASP-7-mediated disease such as an autoimmune disease
XX caused or exacerbated by increased activity of T helper cells. Autoimmune
XX diseases which can be treated using (C) include multiple sclerosis,
XX juvenile diabetes and rheumatoid arthritis. (I) is useful for treating
XX toxemia or pregnancy induced hypertension, pruritic urticarial papules
XX and Rh incompatibility. (I) is also useful as a diagnostic reagent for
XX immune and other disorders, since diseases characterized by
XX overproduction or depletion of lymphocytes in blood or other organs may
XX be detected by monitoring the level of (I) or its mRNA. CLASP-7 has been
XX mapped to the chromosomal location 19q13.2. The present sequence
XX represents the preliminary human CLASP-7 encoding cDNA sequence which is
XX given in the present invention.
XX
XX Sequence 2148 BP; 468 A; 649 C; 625 G; 406 T; 0 other;

```

Query Match 33.6%; Score 2140; DB 22; Length 2148;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2143; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

Qy 4225 cttcagaagcccggaagagctcttgaggcagctgcgaaggtgtgtgcttaagcctg 4284
Db 1 cttcagaagcccggaagagctcttgaggcagctgcgaaggtgtgtgcttaagcctg 60
Qy 4285 ggcacgtcccgagagctctcttcgacatgacctggccaccagagagcctctgtg 4344

```

[illegible]

QY	5425	tcttaacctgvggcaagttccaaagcttgatctcaaaaaggtcctaatactcaagttcaagat	5485
Db	1201	tcttaacctgvggcaagttccaaagcttgatctcaaaaaggtcctaatactcaagttcaagat	1260
QY	5485	gtggaacgtatactttgatactaaagagctcaaaagacgggtgaaactattaaacggaaac	5544
Db	1261	gtggaacgtatactttgatactaaagagctcaaaagacgggtgaaactattaaacggaaac	1320
QY	5545	tatggtcttcgacaatacttcgtttctgcagacggttcaagcgcggatgvgcgcgacaaggg	5604
Db	1321	tatggtcttcgacaatacttcgtttctgcagacggttcaagcgcggatgvgcgcgacaaggg	1380
QY	5605	gagctgcggcgagcaacaacaagcgttaagacgctgcttaagaaacggacaacgctctccctac	5666
Db	1381	gagctgcggcgagcaacaacaagcgttaagacgctgcttaagaaacggacaacgctctccctac	1440
QY	5665	atcaagaacttcgacatacctgtgtctgcacacggagagaaacggtgctgcagccagttgaaagt	5724
Db	1441	atcaagaacttcgacatacctgtgtgtgcacacggagagaaacggtgctgcagccagttgaaagt	1500
QY	5755	gccaatcgagagacatgcaagaaagaaacacgvggagctgagcctttgcacacgagacagacca	5784
Db	1501	gccaatcgagagacatgcaagaaagaaacacgvggagctgagcctttgcacacgagacagacca	1560
QY	5785	ccaagatgttaagatgatacgaagatgvggttccagggctctgtatagvgccacaacggtgaacag	5844
Db	1551	ccaagatgttaagatgatacgaagatgvggttccagggctctgtatagvgccacaacggtgaacag	1620
QY	5845	ggtcccccgtgagvgctggtcccaacggtgtttttatgacgaatacccggaagaccacaagctctc	5904
Db	1621	ggtcccccgtgagvggtggtcccaacggtgtttttatgacgaatacccggaagaccacaagctctc	1680
QY	5905	cgggcaatcaacaacaattgcggtctctgtcttaagagactctctgcagaagaaatgtgaaagatgcg	5964
Db	1681	cgggcaatcaacaacaattgcggtctctgtcttaagagactctctgcagaagaaatgtgaaagatgcg	1740
QY	5965	ctgvggagaaaaaaataagggccctgatttggtgcggagacaaagaaagataccaaacggtgaagttcag	6024
Db	1741	ctgvggagaaaaaaataagggccctgatttggtgcggagacaaagaaagataccaaacggtgaagttcag	1800
QY	6025	cgaacaactacacccgcctgcggtgcggagagctctgcacaacccccctcttaacaaagcgcctgcgccaag	6084
Db	1801	cgaacaactacacccgcctgcggtgcggagagctctgcicagacccccctcttaacaaagcgcctgcgccaag	1860
QY	6085	ctgatggaacccacaccccaacccgggcctcaaggaactccttgaaacagagcaagtttcggaaag	6144
Db	1861	ctgatggaacccacaccccaacccgggcctcaaggaactcctctgaaacagagcaagtttcggaaag	1920
QY	6145	gcaaaactcttgagggcccaacaagggacaaagctgtaaccttaagaaacaagacacggggcctc	6204
Db	1921	gcaaaactcttgagggcccaacaagggacaaagctgtaaccttaagaaacaagacacggggcctc	1960
QY	6205	agcgtcgtctgctgcgagggagatctgcgcccctgtgcacacgtgcggctgvtgvggtgagacaca	6264
Db	1981	agcgtcgtctgctgcgagggagatctgcgcccctgtgcacacgtgcggctgvtgvggtgagacaca	2040
QY	6265	ctgtaacttvggggtcgggcctctgcgcccctgtgtctcccaatctgtgtgcacactgaatgttctc	6324
Db	2041	ctgtaacttvggggtcgggcctctgcgcccctgtgtctcccaatctgtgtgcacactgaatgttctc	2100
QY	6325	cccttttctaatttaaaatggtttttataaagcaaaaaa	6372
Db	2101	cccttttctaatttaaaatggtttttataaagcaaaaaa	2148

RESULT	3
AAH41934	
ID	AAH41934 standard; CDNA; 6828 BP
XX	
AC	
XX	AAH41934;
XX	
DT	30-AUG-2001 (first entry)

DT 30-AUG-2001 (first entry)

XX DE Human CLASP-3 cDNA sequence SEQ ID NO:1.  
XX XX  
XX KW Human; cadherin-like asymmetry protein; CLASP; CLASP-3; immune response;  
KW cell surface molecule; transmembrane protein; immunosuppressive; vaccine;  
KW antiinflammatory; antiarthritic; antianaemic; dermatological; uropathic;  
KW ophthalmological; antirheumatic; nephrotropic; antidiabetic;  
KW neuroprotective; antiashmatic; antibacterial; antisense therapy;  
KW gene therapy; chromosome 1p31.1; ss.  
XX OS Homo sapiens.  
XX XX  
XX PN WO200142297-A2.  
XX XX  
XX PD 14-JUN-2001.  
XX XX  
XX PF 13-DEC-2000; 2000WO-US34171.  
XX XX  
XX PR 13-DEC-1999; 99US-0170453.  
XX PR 14-JAN-2000; 2000US-0176195.  
XX PR 14-FEB-2000; 2000US-0182296.  
XX PR 11-APR-2000; 2000US-0196267.  
XX PR 11-APR-2000; 2000US-0196460.  
XX PR 11-APR-2000; 2000US-0196527.  
XX PR 11-APR-2000; 2000US-0547276.  
XX PR 13-OCT-2000; 2000US-0240503.  
XX PR 13-OCT-2000; 2000US-0240508.  
XX XX  
XX PA (ARBO-) ARBOR VITAE CORP.  
XX XX  
XX PI Lu P, Gorman JD, Candia AF;  
XX DR WPI: 2001-375003/39.  
XX DR P-PSDB: AAB99495.  
XX XX  
XX PT Novel isolated cadherin-like asymmetry protein (CLASP) -3, useful for  
XX PT useful in preventing or treating a CLASP-3-mediated disease e.g.  
XX PT autoimmune disease -  
XX PS  
XX PS Claim 1; Fig 6; 189pp; English.  
XX XX  
XX CC The present invention describes an isolated polypeptide (I) comprising a  
XX CC nucleotide (nt) sequence (Sta) that has at least 90% identity to the  
XX CC sequence given in AHA41934 and is immunologically cross-reactive with  
XX CC the derived amino acid (aa) sequence (SId) given in AAB99495 or shares a  
XX CC biological function with native CLASP-3. (I) has immunosuppressive,  
XX CC antiinflammatory, antiarthritic, antianaemic, dermatological, uropathic,  
XX CC ophthalmological, antirheumatic, nephrotropic, antidiabetic,  
XX CC neuroprotective, antiashmatic and antibacterial activities, and can be  
XX CC used in antisense therapy, vaccine production and gene therapy. CLASP-3  
XX CC related sequences can be used in preventing or treating a CLASP-3  
XX CC mediated disease, preferably an autoimmune disease by inhibiting an  
XX CC immune response. The autoimmune disease is caused or exacerbated by  
XX CC increased activity of T helper 1 lymphocytes (TH1). AHA41893 to AHA41953  
XX CC and AAB99491 to AAB99507 represent sequences which are used in the  
XX CC exemplification of the present invention. CLASP-3 is localised in the  
XX CC chromosome location 1p31.1.  
XX XX  
SQ Sequence 6828 BP; 2082 A; 1384 C; 1432 G; 1930 T; 0 other;

Query Match 33.5%; Score 2132.8; DB 22; Length 6828;  
Best Local Similarity 60.4%; Pred. No. 0;  
Matches 3863; Conservative 0; Mismatches 2262; Indels 273; Gaps 10;

OY 23 ccgaagcgcgcgccttcgcgcacaagatacaagagcgttgcgcgaagcggtgcggaagc 82  
DB 26 ccgagcgccgcgccttcgcgcacaagatacaagagcgttgcgcgaagcggtgcggaagc 85  
OY 83 aggtgtccgcggaagcgttgcgcctcccccactccagcagcggtcgttcgacacatc----- 137  
DB 86 agtctccgcgacaataagtggtctctccccaactgctcaaaaaccttaatatgtgtgga 145

OY 138 -----cctgggggtcccaactgactgaagtgttcgagcccttgacttgagg 184  
DB 146 atatacccatcaacacacacagtgcccttaccgaagacagtagatccagtgatttggaa 205  
OY 185 atgtaactctgagccgaccacagatctcgtgacccggccctccgaaggaacttgaaat 244  
DB 206 ataccatcaatcaacaccccttgctggtgaattctggtgaccttaaggaatttgatgac 265  
OY 245 tcccaagctgaatgacttgagctgtgtcgtgcagcccggaatgctgcgacacaggaagccg 304  
DB 266 ttccctccagatgatattgaaagtgtttatagtcctcgtgactggaagacctgtttcag 325  
OY 305 ggatcccaagatgaataaactgtagtcccaagtgtagggccggtgagagtataatg 364  
DB 326 ctgtacccgtgaagaaagtgaatgtagtccacatglttagagactgataagaagtatacag 385  
OY 365 aggaactggtcatgtgtccacagaaggtatcgtgctcccgagtgagcagcatagaccccg 424  
DB 386 aagactgggcaatctgtcatcagaataatacatcaaatctgggaacaggaattcaatccacata 445  
OY 425 ccacagacacacagcgggagcgacagaagggcctcccgccaggtctttgagcagagt 484  
DB 446 catagataaacaggaagaagaaagggttgccaaaacaaagtttttgaatctgagt 505  
OY 485 cttctgagacgagaggttcggccctgaggaactcgaatgactccggcggtgctgggct 544  
DB 506 aagctccagatggaacacagctacacagatgatacagaatgacttaaaagcgttcaatgt 565  
OY 545 ccccggaagacacccctcgtgaagcagtggtgctctagcatcttgactgaactgaagacctg 604  
DB 566 caatagatgataccccaagggttagctgtggcgttagataltcttgacttgaataatcacc 625  
OY 605 cagctgactcatctgtgcctcctctgtcagaagcgggcgccccaagagatgtgagccgcg 664  
DB 626 ttccgtgactgttcttcccaattacttgatgtaactccaatgaaagaatagacgcgtc 685  
OY 665 gaaatgaaccccttcgacgagcagccggcccgccgtgtcaacctaccgcggac 724  
DB 686 agaattgacccaaggaagaaatcaaacgctcacaaagaacttttgccttgcattcatcac 745  
OY 725 ctgacgaagatgaagcgttgaacgctgtgaacgctgaacgcggcagaccccggaagacttg 784  
DB 746 cagatgaagaagaacaaatagaacggttgaagttctctgatataccacaagaatatttg 805  
OY 785 gacaaagatcttgtaacagtgtcgtgctcaagtctcaagtctgaatgaatgtgagcccatc 844  
DB 806 gtcaaaagcttctgttaaaatgcttatactcaactcaagttgaattgaattgaaccattt 865  
OY 845 ttgggacttggctcctgtatagtgtgcgggagaaagaagaatctcgagaacttact 904  
DB 866 ttgcaagtttgccttatatgatgtlcaaggaaagaagaatctcagaaactttat 925  
OY 905 tgcgcctgaacttggaacccaatgaagggcgtgcttggtgcctatggaacccaaccttcca 964  
DB 926 ttgaccttaattcgaagcagatgaaggggtgtgtacgtccacatgataccaccgtgtgcca 985  
OY 965 tctcaccccttgccgcgtctgcacatctctctgtgactaaccctccacgacatctcc 1024  
DB 986 ttactacccttgcaagatcagcaaatcttttctactacttatctctcccaagatgttttc 1045  
OY 1025 tgttcataagtttgagagaagtgcttcaagcaaggagacatcagtgaagtgttgagcctt 1084  
DB 1046 ttgtataataagctagaagaagtlcctacagcagaagacatcttgagagttgtcagaaacat 1105  
OY 1085 acatggtgttgaagaagttgacacagcgaagaaacaaagaagaactggaagaagtgtggcc 1144  
DB 1106 atagatlttcaagaagacagatgcaaccaagaataagaanaaacttggaaataatgaga 1165  
OY 1145 tggcgcccggaagttctgcaaccccgcttggccgtctacgcgcatgccccttgccttgagcg 1204  
DB 1166 gtcaagcagatcagtttgcgaagaagacttgggaatatatcgcatgctcctttgttgtagctg 1225

QY	1205	ccggagcacttggcccaacatcgttgaagcagcgtctgggcagctctggacccggactc	-----	1257
Db	1226	caatccactttaaataatcgttttagagatctctggagctcttggaaagaaagattctacagaag		1285
QY	1258	-----gactcggagggcgagcgccggccggccctggacacgaacccggcc-----		1300
Db	1286	tagaataatcagtaactcggagaaacgaaaagtgctcttggtccaaaggggaggaattctagatg		1345
QY	1301	---gtcggggggccccagagacccgggcgagttggtgggaacgacgtctctcgtc		1357
Db	1346	ctggcagaacgatactcttgaanaagacaagaagtggagatgatctgttaacttgcagagct		1405
QY	1358	ctcgtccagccagctctaacgtctacacgtctacaaactctcttaagaaagagctctgaagcactg		1417
Db	1406	ctgcacacagcttaactctacagctgacaaattctttaaagcagaagagaaacgcgtctaaagt		1465
QY	1418	acgagagacctcttcaagtcttcctgctgcagacatggagggccggctccgcctgcgcggcac		1477
Db	1466	atgaaagatctcttaacaaattctcttgctgcgatatggaaagccactctctgtcttaagcgac		1525
QY	1478	taagctcgttgaactgcacagctacaaatcgaacattctctcggctctctgaaatccccact		1537
Db	1526	taagacctattacagctcagctcagatcagaatagacattctctccgcacactgaaatccccact		1585
QY	1538	ctgctccctccctcgagctctgctctacataacgaagccctacccggaaccccaagggcccga		1597
Db	1586	atgcgccaactccggagctctgctcttcaagctgaaggtttacactctacagataagattgaaccta		1645
QY	1598	ccaaaggagatctctgagttcccccggccggcgaaagctcatgccccccatcacagctacaga		1657
Db	1646	ccagagaatactctagagttcccccgaaggagatgttatgttttccaaacactactacagaa		1705
QY	1658	acctgctgtacgtgtlacccgcgacagccctcaacttccagcaagccggacaggctccgtgcga		1717
Db	1706	atctctctcaacataaccccctacagacttcaatttgcacatcgttcaagttctgttagaa		1765
QY	1718	acctctgtctggcaggtgacgtatataatgacggcgaggaaccccgaagctctgcggcgca		1777
Db	1766	ataataacagaaagttccagtttatgtatgtagagagatcttcaagaacatggtccatgcccgttaa		1825
QY	1778	tctcttggcaacgtccagctgtgacgtgaattaccccgcgagcgcttccacacgggtgctcac		1837
Db	1826	tctcttggtaatactcagctcgtgttcagaaatttccaagaaagcctatcacgcgtatgatactc		1885
QY	1838	ataacagctcccccggagttctacagaggagttcaagctgcatcttccagcctgtcgtacaag		1897
Db	1886	ataacagagctctccgtgatttctacgtgaagaatcaaggttaagttctctgtactttaaactg		1945
QY	1898	agaaacctaaccctgcgtcttcaactcttaccatgctcagctgcgacgcccggccggcgactg		1957
Db	1946	acacatacactctgcttcttaacttcttaccatgctttagatgtgtacaacaaaataaactctc		2005
QY	1958	ccctggagaaacccggctggggtcttaacttgaatcccaactgcttgagacagggcgccctgaga		2017
Db	2006	ctcttgaacacaacagttgtgatacactgtgataccaatgcttcaagaatgagacggtttgaga		2065
QY	2018	ccggacccctctctgtctccacagtctcgtgtggacacagccggcccgacagctactcgtctca		2077
Db	2066	ctggccaggttttgcttgcgcagctctcaatttggaaaacaccccaagctctattctgtactgt		2125
QY	2078	caaccgactgtgcgctctccgggcatgagcgtctgggtggacggtcacaaggcggttctcagt		2137
Db	2126	ctctcgtgagcttctcttaccactgctgtaaatgttgtagataatcaacaagtgttttaa		2185
QY	2138	tggaggtccaaagccggtctctcgtgcacccccaggaacccctactctgaaacaaattctca		2197
Db	2186	ctggaagtctgttctgttctgcatacacaacaagatccttatcttacaacatttctg		2245
QY	2198	ccctgtgacacggctccggaggaaggagccttccaatctccggtccaaagaaactgtgtcga		2257
Db	2246	ctctcgtgcaaatgtctcgtgatagtgaacacactgttccccgctccgaatttgggacatactgcgaatca		2305
QY	2258	gcgagagggcaacgttgcagcaggagcttgcgggccaagctcttgcagacatgagccttgcggc		2317

Db	2306	tggaataataactctagtaanaatgaaattgaaagagacgactattctcagacatgaaattcaatccagc	2366
Qy	2318	ccgaacccctctgtgacctctccccaacagtgctcggagcaagctcggtctggtcatca	2377
Db	2366	tggaaaccagtggtccggaattcttcattctctctgctcagagtaaacgtaacttttagttat	2422
Qy	2378	ggccccgcatcaatcagtgagccagattgtaaacctgggcgtgtagagccttgaaagcaatg	2437
Db	2426	gacctccctgcatctgctggccaataagtaaacctaggttcaagcatctcttgaaagcaatg	2485
Qy	2438	cccatgtagtgacgcttggtttaccgggagcttgtagagagagcccgaggtgcccggctac	2497
Db	2486	catcaattataaatgacatgacattcaacaaacttggaaagaaatcatgaccagactgycagaa	2545
Qy	2498	gcccaacagctggtctgctcaagttccacacagccttctgcctcttcgtgacatgycagagc	2557
Db	2546	acaagcctctctgcatcaatataattcaattatgtttcttcgcctaccacaatacttaacctaat	2607
Qy	2558	tcccggg-----tggggccccccacgtgacagtgcagagtgycagactgycgcggtg	2608
Db	2606	catctcaaccaggtccctgggggttttgtagagatcgaagtgcattatgycacatactgctagat	2656
Qy	2609	gctctggtctgcgcgcgcgcgaagcctctactctgctgcggttccaaagagatcatcagcagcgaac	2666
Db	2666	ctgcggtgtagagcctctgcaagccttaatttaaatcggtctgcgaagccttagtaataagatc	2722
Qy	2669	ctgacactgcgcgttgccccctggcgccttggtgattagagaggt	2705
Db	2726	cagataatctcggagactccacgcttcacagcaatgatagtctgatacatcagcggaat	2785
Qy	2710	-----	2705
Db	2786	agggtttagatcgtccaatccctctggtttaacactggtgtccaaaagctgcccacatg	2845
Qy	2710	-----	2705
Db	2846	gattcaaccccagctcagaatgacgaatcaacacagagctatgtagtcgaagtgttaatcgt	2905
Qy	2710	-----tcccgcacatctgggcca	2722
Db	2906	tgtcttcgcacaacagagagatgcaagttcttcttaacaacataaagggagcgttaaccaata	2966
Qy	2726	gcaagctgctcagcagagagcctggtctctcagtgaggtggtgtcagcagctgacgtacag	2785
Db	2966	aaaagctttctcacagagagcctggtcttcttcagtgaggtgtttgttcagtggtcagcgtctcgg	3025
Qy	2786	aggccaactccagcagcgcctgtgtctcttctccaagctcaatggtgaaagatagtcgcgtc	2845
Db	3026	aatcagcttctgcaacaaagcctggtctcttcttgaaatagtgtaagagacatggtgcac	3085
Qy	2846	acctgctgtggcagcagcagtaagacaccccgacagcctgcgtctcccgcggagcctcc	2905
Db	3086	atttaacttaatgataaataattgaggctcccaagaaagtcgttttccagaagactttca	3145
Qy	2906	tggagacatacatgctctgtgtgggtcgtgtgtagccttgtaggttcatcacccgtgtgcaca	2965
Db	3146	tggatgacattgcagcgtcttgtagcagcaggttgcagtgatagttatagttcaagattcaga	3205
Qy	2966	aggatgtgtagagctggtccgagcacctcaacgcagcagcctggtcttctctcactagtgacctc	3025
Db	3206	aggacaagaaatggttgagagacatcaatacaagccttgatcttcttccaatgactgt	3265
Qy	3026	tgtccctgtgtgaccggggtctgtgtctcagctgtgtccgggccaactataaagcaggtg	3085
Db	3266	tgtctgtataggaagaggtattgttttggccttataaagttccgtctataaaacaggt	3325
Qy	3086	ccacgcggtcctcagctgtccctcaataatccagcagcctctgtagacctgcgcagtgatca	3145
Db	3326	cttcaagcttactcaattaccgaatccaggtgtcttgggtgtcccttgaggtctggaatttc	3385
Qy	3146	ccgcgactctgtgagcacaagagactagtgtagccctcaacctccctctgtctgccctgt	3205

Db 3386 taagaaatcatctgcagtgaaatgagcaactatgtaataacttaacctgaaccttactcta 3445  
QY 3206 caactcagactgcgcctccctctgtctcctccacacacccctccagagctccacctct 3265  
Db 3446 ctccacctgactccatccactctctctctctgcaacatccatccagagttccgattt 3505  
QY 3266 ccagccaaagcccgagccccaaggtgacacagatgttccgaactgagtgacatccacgc 3325  
Db 3506 ctacgaatgttaacagaccaaagaatgtgcaaatatgtttcaatatccgtcccttcgcgc 3565  
QY 3326 agacagacttccactggtgctcctgctgacgagcttggaactggaactgcccgaacccgag 3385  
Db 3566 aacagaaatatttggcagagactgtgttaacgagctggtctgcatatttagaccctgag 3625  
QY 3386 ctgaagggagcatctcgtgtgacaaagaagccaatcagtgctgtgacagccgtgtatgt 3445  
Db 3626 ctgaagagactgttggattgataaagaagtaataatgtataatgtatacaatttactctca 3685  
QY 3446 gcaatgaacatgaaccccggtacgccgagggccactgtgaaggtctgtgtgcccgaactgt 3505  
Db 3686 gtacagactcagaccccggtactcctgcgaacctcagaataaaggtccgagtgccactgtgt 3745  
QY 3506 acctgcgaactgtcttcgatgtcagggatccacttgccagcgctgacatgtactgtcgtgag 3565  
Db 3746 atctacctgattgtgtattatcatgtgaactgttacctgaactgtgtatgttttaacgaa 3805  
QY 3566 gccacagtgacgctgtaagaactgtgcccataatgcttgaactcagaacagaaagcgaaagg 3625  
Db 3806 ctacacatacaacaggaagaac-----caattgtataccactgtatgtatlaagaagtgt 3859  
QY 3626 acattgcgggtatccatcaaacccctctgtgccaatgccaatgtctgtgtgccccctgaagcc 3685  
Db 3860 aggcgcggaagtatgataaagccagaccgtgtgcacatgcgaatgcgaagggacatcgttcccc 3919  
QY 3686 ctggtctccggccagcaatctcccaaggggccacaacggtctctccgcgacaggtgtgccc 3745  
Db 3920 aactaacaaggtcgtgcgaattctctcctccacgtcacaagagtgccaggaacacactact 3979  
QY 3746 tctctgtgtagttaaagccggaacttgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3805  
Db 3980 ttccacgaagaataacatgcgaagcccttctgtaactgttgggttctccaaaaatgacga 4039  
QY 3806 agccgcgctcctcctgcagcgctgtgcccactgtgacactccccccagccgtggagcgtgtgt 3865  
Db 4040 atgaacaagttctacagaagtggttttaacagatctctcagttcttcgcagctaaacgcgtat 4099  
QY 3866 tggactgtgtacacttctgtgccttgtgactgtgatacaaggggaaaaagcccttgaac 3925  
Db 4100 taagatcgtcttactctgt 4159  
QY 3926 gcatcaacagcctcaatctcaaaaaaattctctgatatgaagggcgctgagggaaagcca 3985  
Db 4160 gaattgaataagcttgacctttaaagaatacaaaaatcagagaaagacttgaagaagta 4219  
QY 3986 ttctgtgttaacatcggagctcgaacagaatgtgttcggcgaaagtctgt----- 4033  
Db 4220 ttcttggggacataggtgtccagggcaagaatgtgtacggcgaaagccgaggaacagcttcgaa 4279  
QY 4034 -----agaggaagccggttgggaatccagagaacgtgcggtgcggaagagcgttca 4084  
Db 4280 gaagccatctggaagtgtccttggagtaagaataatttggaggtgaggaagaatattga 4339  
QY 4085 caacatggaaagcaactaagaccggtgtgataaagccaagatgaatgtgaaacagagag 4144  
Db 4340 ctcaacgtgcgtcaaaaacagagaagcttgcaaatcaagaacagaaagatttgaacagag 4399  
QY 4145 ccttgtgtgaaggaacctgtgcacacgaggaagccactgagtggttctgcgacacactgaa 4204  
Db 4400 caactgtgtatgtaaaacctgtgcacagaaagcaacctaaatcaatttaagtaatactagaa 4459  
QY 4205 tcatctgtcagagagctgtgacttcaagaagcccggaagagcgtcttggggcagtgctga 4264  
Db 4460 ttgttgtacagacgttctgttaacggaatccaagaagagcattctgtgtgagtgctaa 4519

QY 4265 aggtgtgtctgtacagcctgggcaggtgcccagagtgccctctcttctgcaagatggcctgtg 4324  
Db 4520 aagtgtactcaacagcactggtgttaaccaaagtgcagtttactatcaacaacgttcttg 4579  
QY 4325 ccaccacagagggcccttgtgtccaaagttcccgagcgtgtgttctgagaggaagacagagc 4384  
Db 4580 ctacacagagagcccttgtgttccaaagttccctgaactcttatttgaagaagaagacagagc 4639  
QY 4385 tgtgtgcgaacctgtgtcctgtgagccctcagacactgtgtgagccagccatccagccatcc 4444  
Db 4640 agtgtgcgatttaatgactcaggtcttcgcgaactgtgagcagatcagatcgtgtcaatac 4699  
QY 4445 gcaacgcagcagcgccctgcgtcgtatcctgtcctatgtgcagacagaacttcgagatcgcgaca 4504  
Db 4700 ggtcacaccccagctgtccctcccttaccataatgagcaaaacttvgagtgtggataa 4759  
QY 4505 actttgcgcggtgtgaagtgagtgacagatcctctctgtccctgtgtgtggagcagagc 4554  
Db 4760 actttgcgaaggttaaaatgcaaggtacaaatgtacatactccctctgtgtgtgcacatctc 4819  
QY 4565 agaaactcagtgaaagagacactgtgcagcgttccactcaaaacccatccactatgctgag 4624  
Db 4820 agaatttaatgaagaatctcttaagacgtctctcctaagacataattgacatattgtctgaag 4879  
QY 4625 aggacatvgggcgtcgcggagacacacttcgcagagcaggttccaggaactgtatcaac 4684  
Db 4880 aagatctggaattgaggaagaaacaaacttccctgacaggtccaggaatcgtgtttccatc 4939  
QY 4685 tgcacatgatccctgcagacacggtgtgaagatgtgaaggaacacacagaggaacccctgagtg 4744  
Db 4940 tccatgatctctctctgtatctgtgaatgtgaaggaacacacagaggaacccctgagtg 4939  
QY 4745 tcatcgacctatgtacagaaattgcgccgggtctacacaggtctccacggacttgcgtga 4804  
Db 5000 tgattgtactaatgtacagaattggccaaggtttacagaaactctccaga---ggcatgtga 5056  
QY 4805 cctgtgtgcagaaactgtgcgggaagcagcggtgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4864  
Db 5057 cctgtgtgcagaaactgtgcgggaagcagcctcaggaacgaagaacatcactgtgaactgtcac 5116  
QY 4865 agtgcattgtgcagcgccgcgcctcgt 4924  
Db 5117 agtgcattgtgcacacacagcagcactgt 5176  
QY 4925 gccacctgcggt 4984  
Db 5177 aatacttccgt 5236  
QY 4985 ccgcacatctccgaagacactcctgt 5044  
Db 5237 ctgcggtctccagatgtatgtgtatctccagatgtgaagatctgtccttgaataact 5296  
QY 5045 tcaactgcgt 5104  
Db 5297 ttacagatcagagactgt 5356  
QY 5105 ggtcttcagagcggtgtgaatgtatgt 5164  
Db 5357 gcatgtatgaagaagtttaagtttaagaagtttaagaagtttaagaagtttaagaagtttaagaagtt 5416  
QY 5165 gtgatacaagaagcgtgcgcggtgtgcacggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5224  
Db 5417 gggatgcaaaagaactcccaatltcatgttlaaacttcaagaagcatlcaagaanaattg 5476  
QY 5225 tgcacagaagttccgcgtgcggagcggtgttcggagacgtatttccgcgtgtgtgtgtgtgtgtgt 5284  
Db 5477 ttcatcagagtaactgctgcggagcggt 5356  
QY 5285 ggcgccacttcggtgtgaactgtgaatgtgaagagagttgtgtataaggaacatcgatacga 5344  
Db 5537 gaaccaagttcgggatttggattgaacaagaattgttatacaagagcctgcgcaataacga 5596



QY	5345	agctcgacagagatctccaacccgctctgagagagcttccaaaccggaagagatttgcagccgcg	5408
Db	5597	aactctgacagatatactccacaagatctgagaggattcttaacggaanaagatttggacagatcg	5656
QY	5405	tgcgttggattatataaagactcttaccocgtgtagacaagccgaagcttgactctacaanaag	5464
Db	5657	tgcttgtagatatacaaaagactcttaaccctgtagacaagctgtaaatatagatcctaacaag	5716
QY	5465	ccatactccagatatacgtatctgagaaocgtaacttgaatacctaagactcaagaccgcg	5524
Db	5717	catatactcagattatcctaactgtagagccatacttcttacacatatagatgaaagacaga	5776
QY	5525	tgacatacttggacccgaactatgvgcttcgcacactccctgtcttcgcaagccgttcaagc	5584
Db	5777	tcacatacttcgaaanaaatatacaatcttcgctcgtacatactcgtacgtlaaaccttactt	5838
QY	5585	cgaatggcgcgcaacaagggagctctgcgcgagcaacacaaagcgtataagacgtctcagca	5644
Db	5837	tagatgvcgcgtgcccacgtgggaactctcatgaaacatactcaaaaggaagacatctcgacta	5896
QY	5645	ccgaacacagccctccctacatacaagaactcgcactccgtctgtgccaccggaagagacg	5704
Db	5897	cgctcactgccttctcctataataaaagaaggtacaaatgtcactcaataaagaagacata	5956
QY	5705	tgctcgacgacgaagctgagctgtagacatcgtaagagacgtgaagaagaacgagacgctc	5764
Db	5957	tcctaaacaacaattgaaagctgtgcattgtgagagacgtgaaagaanaagacacagagcttgc	6016
QY	5765	ttgccacacggaacgaacccaacagatctgactaaagatgcgtaaagaatggcttcacagctcg	5824
Db	6017	ttgcaacaacatacgaagatccgcagaccccaataatgttccagtgtactccagagatctg	6076
QY	5825	tagggcccaacgctgaaacacagagctccctctgagagcttgcccaagctgttttaagcagaagctc	5884
Db	6077	tagggacacacaagctgaaacacagggccctctggaagattgcccagagcttctcgtctgaataac	6136
QY	5885	cggaaagaccccaagctctctccggtacacacaacaatctggctgcctcgtctcaagagactct	5944
Db	6137	ctagtgcaccaaagctctctcagacatacataataactgtagactcgtcttaagaattta	6196
QY	5945	gcaagaaatcttgagatctgcgtcgcggaanaataaagccctgatttggcccggaaccgaag	6004
Db	6197	ctaaagaatctgtaagatgctcttaagaanaataaagacttaatttggccggttcaaaag	6256
QY	6005	agttccacacgctgagctgtagcgcgaactactgcgcgccttgaggagagctcttgacagccccc	6066
Db	6257	agttacaaagggatctg--gggaactatacttcgctcttaagaagagccctcaagccctaga	6314
QY	6065	ttaacccgcgcctgcgcccaagctgattgagacccaccccc--acccgcgcctgaagacac	6120
Db	6315	tcacagaagaatctccctcaagctatacccaagccagatattgcttgcctccctgcacaaagatctc	6374
QY	6121	ttgaacagagcaagcttccgaaagaagcagaactcttgagc	6158
Db	6375	ttcagtcgaatgagcttctgcgaanaatgagctctcttaaac	6412

RESULT	4
AA507373	
ID	AA507373 standard; cDNA; 7215 BP.
XX	
AC	AA507373;
XX	
DT	26-SEP-2001 (first entry)
XX	
DE	Human cDNA encoding CLASP-5.
XX	
KW	Human; CLASP-5; cadherin-like asymmetry protein; immune gateway;
KW	immunogen; antibody; autoimmune disease; rheumatoid arthritis;
KW	multiple sclerosis; leukemia; insulin dependent diabetes mellitus;
KW	acquired immunodeficiency syndrome; AIDS; ss.
XX	
OS	Homo sapiens.

XX	Key	Location/Qualifiers
EH	CDS	112..6048
FT		/*tag= a
FT		/product= "CLASP-5"
FT		1498..1519
FT	primer_bind	/*tag= b
FT		/note= "PCR primer HCS951"
FT	primer_bind	3087..3113
FT		/*tag= c
FT		/note= "Antisense oligonucleotide AAS07384"
FT	primer_bind	3087..3113
FT		/*tag= d
FT		/note= "Antisense oligonucleotide AAS07384"
FT	primer_bind	3232..3259
FT		/*tag= e
FT		/note= "Antisense oligonucleotide AAS07384"
FT	primer_bind	complement (3642..3660)
FT		/*tag= f
FT		/note= "PCR primer HCSA510b"
FT	primer_bind	4728..4756
FT		/*tag= g
FT		/note= "Antisense oligonucleotide AAS07384"
PN	WO200142296-A2.	
XX	14-JUN-2001.	
XX	13-DEC-2000; 2000WO-US34163.	
XX	13-DEC-1999; 99US-0170453.	
PR	14-JAN-2000; 2000US-0176195.	
PR	14-FEB-2000; 2000US-0182296.	
PR	11-APR-2000; 2000US-0196267.	
PR	11-APR-2000; 2000US-0196460.	
PR	11-APR-2000; 2000US-0196527.	
PR	11-APR-2000; 2000US-0196528.	
PR	11-APR-2000; 2000US-0547276.	
PR	13-OCT-2000; 2000US-0240503.	
PR	13-OCT-2000; 2000US-0240508.	
PR	13-OCT-2000; 2000US-0240543.	
PR	13-OCT-2000; 2000US-0240539.	
PA	(ARB0-) ARBOR VITA CORP.	
PI	Lu P, Garman JD, Candia AF;	
XX	WPI: 2001-367865/38.	
DR	P-PSDB: AAU04024.	
XX	CLASP-5 polynucleotides, proteins and antibodies are used to prevent or	
PT	treat a CLASP-5 mediated disease, such as an autoimmune disease e.g.	
PT	Rheumatoid arthritis -	
PS	Claim 1; Fig 6; 188pp; English.	
XX	The sequence encodes a cadherin-like asymmetry protein, CLASP-5, which	
CC	is a transmembrane protein of the immune system involved in the formation	
CC	of the immune gateway. CLASP-5, polynucleotides encoding it and an	
CC	anti-CLASP-5 antibody are used to prevent or treat a CLASP-5 mediated	
CC	disease, such as an autoimmune disease caused or exacerbated by	
CC	increased activity of TH1 cells. These diseases may include Addison's	
CC	disease, rheumatoid arthritis, dermatitis, Multiple Sclerosis, Systemic	
CC	Lupus Erythematosus and Autoimmune Thyroiditis, Inflammatory conditions	
CC	(e.g. ischaemia-reperfusion) and responses, Leukaemia, acquired	
CC	immunodeficiency syndrome, AIDS, syphilis, typhoid, Burkitt's lymphoma,	
CC	Grave's disease and insulin dependent diabetes mellitus. CLASP-5 is used	
CC	for inhibiting an immune response in a cell (T cell or B cell) by	
CC	interfering with the expression of a CLASP-5 gene in the cell, the	
CC	ability of a CLASP-5 protein to bind to another cell or the ability of a	
CC	CLASP-5 protein to bind to another protein. The antibody binding CLASP-5	
CC	is used to inhibit an immune response is a subject. The polynucleotide is	
CC	used to detect CLASP-5 expression in cells and for diagnosis of	







[illegible]

QY	5391	atttggcgcagcagctgcgtttgagatataccaagactcttaaccgtttgcagacaagtc	5450
Db	5367	ttttgttcagaaatttgggaagtgatataaagactccactctctgtgacaacaagaatt	5426
QY	5451	tgacctcaaaaaggcctactacatccagatatacgttatgttggaacccgactttataccta	5510
Db	5427	ggatctcaacaagagcctacatacagataactctgtgtgagccctactcttgatgattga	5486
QY	5511	gtctcaagagccgggtgagactactctttgacccgcaactatgvggcttcgcacatctgtctg	5570
Db	5487	gatgaagaacagcgggtacacataacttttggaagaattcaactcccgsgggttcaatgatac	5546
QY	5571	caagcgcgttcacgcgcggaatgvgcgcgcacacagvggagctgcccagacaacaagtgtaa	5630
Db	5547	caaccgcgttcacccctgtaggggcgcgcctcgcgggagagctgcatgagcagatagagaagaa	5606
QY	5631	gacgctgtctcgacacccgacacagcctctccctataataagctcgcgtccgtgtgtgcca	5690
Db	5607	cacagctctgcagccacatagcatgacgcctctccctacatacaagaagccagatccagctacca	5666
QY	5691	ccggagagagagacagcgtgctgcacgcgcagctgaggtgagctgcatcagagaagacatgc	5750
Db	5667	gaagagagagattgtgttttgacaacccgatgaaagtgtccattgaaagacatgaaagaagac	5726
QY	5751	acggagagctggtcctttgcacccgcgcagagaccacacagatgctaaagtctacagatggt	5810
Db	5727	cctgcagtttagcagtttgcctacttaacacagagcgcgcgtatgacaagaattgcttcagatggt	5786
QY	5811	gcttaagagcctctgtgaaagggcccaacccgtgaacaaaggttcaccttgaaagttgcccaggtgt	5870
Db	5787	gtctgaagagcctctgtggaagcttactgttaatacaagggacacattggaatgaaccacagttgt	5846
QY	5871	tttaacagagatccctgcggagaagaccacaaagctcttcgcgatactacaacaacaattgcgcctgt	5930
Db	5847	tttgctgtgaatctcctgtctgataccaagaactctatccagataccaacaagttgaggtatg	5906
QY	5931	cttcaagagactctgcaagaatgtgtgagatgctgcgtcgcgaaaaataagccctgattgg	5990
Db	5907	ctttagaagattcatcacatgagatgtgtgtgaagctgtgaagataaaaacacagctctcatcac	5966
QY	5991	gccgcgacacagaagaggttccaccgtgtgagctgagcgcacactacgcgcgcgtgcgggagc	6050
Db	5967	ggcagacacagagggataatctcagcaggaactcaaaaagactatacaagcctaaagagaa	6026
QY	6051	tctgcagcccccgtcttaccacgcgcgtcccccagctg	6087
Db	6027	cctcagcccaatgatgcagcggaaatctccagaactg	6063
RESULT 5			
AAH99572			
ID	AAH99572 standard; cDNA, 1652 BP.		
AC	AAH99572;		
XX			
DT	16-OCT-2001 (first entry)		
XX			
DE	Human protein encoding cDNA sequence SEQ ID NO:407.		
XX			
KM	Human; cancer; ulcer; HIV infection; human immunodeficiency virus;		
KM	antiinflammatory; antihemmatic; antiarthritic; immunosuppressive;		
KM	antibacterial; endocrine; cardiact; central nervous system; virucide;		
KM	anti-HIV; fungicide; antimitogen; cardiovascular; antianaemic; anaemia;		
KM	antiaegregant; haemostatic; vulmerary; antitumor; osteopathic; eczema;		
KM	dermatological; antiallergic; antisthmatic; antidiabetic; cytostatic;		
KM	neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;		
KM	immunostimulant; gene therapy; antitense therapy; vaccine; inflammation;		
KM	antianaerobic; rheumatoid arthritis; septic shock; pancreatitis;		
KM	cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;		
KM	genetic disease; haematopoietic disorder; platelet disorder; asthma;		
KM	thrombocytopoenia; osteoporosis; severe combined immunodeficiency;		
KM	allergic rhinitis; diabetes; multiple sclerosis; depression;		
KM	Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;		

KW neurological disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN M0200153455-A2.  
XX  
PD 26-JUL-2001.  
XX  
PF 22-DEC-2000; 2000MO-US55017.  
XX  
PR 23-DEC-1999; 99US-0471275.  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
XX  
PA (HYSEQ-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT.  
XX  
DR WPI; 2001-457603/49.  
DR P-SDB; AAM25631.  
XX  
PT Isolated human polynucleotides encoding polypeptides, useful for the  
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
PS  
XX Claim 1; Page 505; 1217pp; English.  
XX  
AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
CC AAM25963. The proteins can have activities based on the tissues and  
CC cells they are expressed in, such as: antinflammatory; antirheumatic;  
CC antirheitic; immunosuppressive; antibacterial; endocrine; cardiac;  
CC central nervous system; Virucide; anti-HIV; fungicide; antimutagen;  
CC cardiovascular; antiaemic; antiagregant; haemostatic; vulnery;  
CC antilacer; osteopaltic; dermatological; antiallergic; antisthmatic;  
CC antiblastic; cyostatic; neuroprotective; antidepressant; nootropic;  
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
CC encoding them can be used in gene therapy, antisense therapy and vaccine  
CC production. The proteins and polynucleotides are useful for screening for  
CC agonists or antagonists of a protein and for the treatment and diagnosis  
CC of disorders associated with the activity of a protein e.g. inflammation,  
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
CC anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,  
CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
CC neurological disorders.  
XX  
XX  
SQ Sequence 1652 BP; 377 A; 478 C; 479 G; 318 T; 0 other;

Query Match 23.0%; Score 1465.2; DB 22; Length 1652;  
Best Local Similarity 99.7%; Pred. No. 9.3e-287;  
Matches 1489; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 4879 gggcgccgctcgctgctgagtaactcgcgcgtctcagagaccagcgccacctg 4938  
DB 2 gggcgccgctcgctgctgagtaactcgcgcgtctcagagaccagcgccacctg 61  
QY 4939 ggtctgcttccttcagaaacatcatcaaaagtgtctagagagtcgcgcacatcttc 4998  
DB 62 ggtctgcttccttcagaaacatcatcaaaagtgtctagagagtcgcgcacatcttc 121  
QY 4999 gaacatctgctcgccagagagaggtctcgtccggaagaaacttaactagctgg 5058  
DB 122 gaacatctgctcgccagagagaggtctcgtccggaagaaacttaactagctgg 181  
QY 5059 ctggttaggggtctctgtaacagcagcggtacttaccatggcggtcttaccagagcg 5118  
DB 182 ctggttaggggtctctgtaacagcagcggtacttaccatggcggtcttaccagagcg 241  
QY 5119 gtgaatgaggtctacaagaactcatcccatcctggaaagcccaacgctgactacaagaag 5178  
|||||

DB 242 gtgaatgaggtctacaagaactcatcccatcctgtgaagcccaacgctgactacaagaag 301  
QY 5179 ctggccgcggtgcaagcgaacaaatctgagagagccttcaccaagatcatgcaacaaagtcc 5228  
DB 302 ctggccgcggtgcaagcgaacaaatctgagagagccttcaccaagatcatgcaacaaagtcc 361  
QY 5239 ggtctggagcggtgtctcggaagcatttcgcgcgtggtgcttcacggcgccacctg 5298  
DB 362 ggtctggagcggtgtctcggaagcatttcgcgcgtggtgcttcacggcgccacctg 421  
QY 5299 gacctgtgagcaagagatttctgtacaagagccatcgatcaacgaagctgagcaagatc 5358  
DB 422 gacctgtgagcaagagatttctgtacaagagccatcgatcaacgaagctgagcaagatc 481  
QY 5359 tcacacgcggtggaagatttctacacggaagatttgcgcgcgcgtctgtgaattatc 5418  
DB 482 tcacacgcggtggaagatttctacacggaagatttgcgcgcgcgtctgtgaattatc 541  
QY 5419 aaagacttcaacccctgtagaagaagtcgaagctgactacacaagaagcctacatcagatc 5478  
DB 542 aaagacttcaacccctgtagaagaagtcgaagctgactacacaagaagcctacatcagatc 601  
QY 5479 acgtatgtggaacgcgtacttgaatacctacagactcaagacgcggtgactacttgac 5538  
DB 602 acgtatgtggaacgcgtacttgaatacctacagactcaagacgcggtgactacttgac 661  
QY 5539 cgcacacttggtcgtcgacatcttcgtcttcacgcgcgcgttcacgcggaatggcgcgca 5598  
DB 662 cgcacacttggtcgtcgacatcttcgtcttcacgcgcgcgttcacgcggaatggcgcgca 721  
QY 5599 caagggagctgcccgaacacacaaagcgttaagacgcgtcctcaagcacagacacgccttc 5658  
DB 722 caagggagctgcccgaacacacaaagcgttaagacgcgtcctcaagcacagacacgccttc 781  
QY 5659 ccctacatcaagactcgcgtacgt 5717  
DB 782 ccctacatcaagactcgcgtacgt 841  
QY 5718 ggaagtgcacatcgagagcatgcgaagaagaacacgcggaagcttgccttcgacagagca 5777  
DB 842 ggaagtgcacatcgagagcatgcgaagaagaacacgcggaagcttgccttcgacagagca 901  
QY 5778 ggaaccacagatgtctaaagtctacagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5837  
DB 902 ggaaccacagatgtctaaagtctacagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 961  
QY 5838 gaaacagaggtccctcggtggt 5897  
DB 962 gaaacagaggtccctcggtggt 1021  
QY 5898 gctcttcgagcatcacacaacattgcggtcgtctgtcttcagaagactctgc-aagaatgtg 5956  
DB 1022 gctcttcgagcatcacacaacattgcggtcgtctgtcttcagaagactctgcagaagaatgtg 1081  
QY 5957 aagatgtgcgtcggaaaaaataaaggccctgattgggcccgaacagaagagatcacacgtg 6016  
DB 1082 aagatgtgcgtcggaaaaaataaaggccctgattgggcccgaacagaagagatcacacgtg 1141  
QY 6017 agcttgagacgaactactgcgcgcgtcggtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 6076  
DB 1142 agcttgagacgaactactgcgcgcgtcggtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1201  
QY 6077 tgcgccagctgattgaccccccaccccgagcctcagaagaactcttgaagagcaagtt 6136  
DB 1202 tgcgccagctgattgaccccccaccccgagcctcagaagaactcttgaagagcaagtt 1261  
QY 6137 tcggaagagcagacactctgagcccaacaagacaaagctgttacctagaagaaacagcacc 6196  
DB 1262 tcggaagagcagacactctgagcccaacaagacaaagctgttacctagaagaaacagcacc 1321  
QY 6197 cgggcctcagctgtctgt 6256  
DB 1322 cgggcctcagctgtctgt 1381



QY	3673	ggccccctagaacctgtgtcccgccagcatctcccaggggccaaccaagcgcttcctgc	3732
Db	1219	acalcgtgtccctcaactaacaagaagcctgtgcgaattctctctcaactcaacagatgtgcag	1278
QY	3733	gcagggctgtgccctctctgtctgtgtcaagccgagacctgtctgcgtgtgtctgtgagtg	3792
Db	1279	caaacactactacttcttcagagaatcaaatgtcgaagcctttgtactgtctactcttgggtc	1338
QY	3793	ctgaanaacacgcgcgcgcctctgcagcgcttggccactgtacatcccccag	3852
Db	1339	ctcaaaaatctcagatgtgaataacagctctacacagaagtgtgtttacagactctctcagcttgag	1398
QY	3853	ctggagacgtctgtgttgagactgtgtactcttcttcctgcgcctgtctgtgtgtataaaggagaa	3912
Db	1399	ctaaacgcgctattatgactctgccttactctctgtgtctgtctgtgtgtatataaaggagaa	1458
QY	3913	aagcgcttgaacgcacalcaaacagcctcacaattcaaaaatctctgtatataagcgcg	3972
Db	1459	aaagtgtttgaacgaatgtaatagactgtacaccttaagaataatcaaaagatcagtgagcaag	1518
QY	3973	ctagaagaaagcactctctgtgtgtatccatccggaagcttcgaacaaagtgttcctgcgcgaagtcgt	4032
Db	1519	cttgaagaagactactctcttgcgagcatagtgctgcacgaagaatgtgtatcggcggaagccga	1578
QY	4033	g-----agaaagacccgcttctgttgagaatccagagaaacgtgcgcgtg	4071
Db	1579	ggacagctctcgagaggaagcccatctgtgaagtgtccttgcgtgaggtcgaagaatttgatggtg	1638
QY	4072	cggaagaagcgctcacacacactgtgaagaacaacctgaagccgcgtgtgacagaaccaagatgaa	4131
Db	1639	aggaaaaagatatgtactactctgtgcgtcaaaaacacgaaagacttgcgaataatcaagagcagag	1698
QY	4132	atggaacacagagcgctctgtgtgtgaagggaacccgtgcacccgagcgagcgaaactgtgtcttg	4191
Db	1699	attgaacacggaagcactgtgatgtgtgaaacctgcgtctacagaaagcaaacctatacttta	1758
QY	4192	gaacactgtgagatcalctgtcagacggtgtgatgtcttcagaagcccggtgagagcgcttg	4251
Db	1759	gactcatlaagatgtgtctgtctgcagaccgttctctgtacagaaatccaaagagagcactctt	1818
QY	4252	ggggcagtgctcgaagaaggtgtgtgcgtctcaagcctggggaggtgcacagagtcgcctctcttg	4311
Db	1819	gggtgagatgtcnaaaagatgctactacaacaagcagcgccgtgaaaccaaagtgagcagttatacta	1878
QY	4312	caggaatgtccctgcgcacccacagagggcgccctgtgtgtccaaattcccgagactgatttcag	4371
Db	1879	caacactgttttctatacaagagagcgcttggtttcaagtcttctcgtaacctcttaattgaa	1938
QY	4372	gaggaacacgagactgtgtgtgcgcgaacctgtgcctgaggtctctctcagcaactgtgtgcagccg	4431
Db	1939	gaagaagacaaagacggtgtgtctgattatgtcctcaggtcttctcgcacactgtatagacgtatgc	1998
QY	4432	atcagcaacatccgaacagcagcagcgccagcgctgtactactgtctatctgcgcagaaatctc	4491
Db	1999	atcgtcatcaataatgcgtacaccccagctgtccctcttaactactatgaatgaagcaaaactt	2058
QY	4492	gagatcgcgcacaacttgtccgctgtgaagatgcagttcacatgttctctctgtccctg	4551
Db	2059	gagatttggataaacttctgtgcaggtttaaatgtcagttacaaatgttaactctctcttg	2118
QY	4552	gttggggacgcgcgaagaactctcagtgtgaagcagacccgtgcagcgttcaactaaacatctctc	4611
Db	2119	gttgggcacatctccgaattttaaigtgaagaattcttaagcgttctctcctaaagactatattg	2178
QY	4612	accatctgtgaagagagacatgtgcgtgcggaggaacagaccttcgagagcagttccagagac	4671
Db	2219	acatattgtcgaagaagatcgtcggaattgtgaaggaacaaacttctccgtatccagttccagat	2238
QY	4672	ctgatgttcaacctgcacatgtactctgtacgcgaacggtgtgaagatgtaaaggaacaccagag	4731
Db	2239	cttgtttctacatctccataatgtacttcttgcgtactgtgtatgtgttaaatataagaaacacagag	2298

[illegible]

```
Db 3379 ctccaggagctctagaccacacagatcaggggccttggaaagtgtgccaggtttt 3438
Qy 5872 tttagcagagatcccggaagaccacagctctccggacacacaacaaattgctgtgc 5931
Db 3439 ctgtctgaataactactgtgtaccacaaagccttcacagacacataaactgtgactctgc 3498
Qy 5932 ttcaaggaactctgcaagaatgtgagatgcgctgcggaanaataaaggccctgattggg 5991
Db 3499 tttaagatttctactaaagtgtagaagtgcttaagaanaataaaggcttaattggg 3558
Qy 5992 cgggaccagaaagagafacacgcgtgagcggtgagcgcaatactgcccgtctgcggaggt 6051
Db 3559 ccggttcaaaagagatcatcaaaaggaaatcg--gggaactatcttcgccttaagaagggcc 3616
Qy 6052 ctgagccctcgtcttaccacagcgctgcgccacgtgatgtgacaccacccc---accgg 6107
Db 3617 ctacagccctagctacacagaaagtcctcagttatcccaagccaagtattgtgtccctcg 3676
Qy 6108 cctcaggaactccttgaacagagcaagttccgaaaggcagactctgagc 6158
Db 3677 ccacagagatctcttcagtcgaatgagcttccgcaaatgagatctcttaac 3727
```

## RESULT 7

AAH41912 ID AAH41912 standard; cDNA; 4144 BP.

AC AAH41912;

DT 30-AUG-2001 (first entry)

DE Human CIASP-3 cDNA sequence Fig 4.

XX Human; cadherin-like asymmetry protein; CIASP; CIASP-3; immune response;  
XX cell surface molecule; transmembrane protein; immunosuppressive; vaccine;  
XX antiinflammatory; antirheumatic; antianemic; dermatological; uropathic;  
XX ophthalmological; antipneumatic; nephrotropic; antithyroid; antidiabetic;  
XX neuroprotective; antisthmatic; antibacterial; antisense therapy;  
XX gene therapy; chromosome 1p31.1; ss.

XX Homo sapiens.

XX PN MO200142297-A2.

XX PD 14-JUN-2001.

XX PF 13-DEC-2000; 2000WO-US34171.

XX PR 13-DEC-1999; 99US-0170453.

XX PR 14-JAN-2000; 2000US-0176195.

XX PR 14-FEB-2000; 2000US-0182296.

XX PR 11-APR-2000; 2000US-0196267.

XX PR 11-APR-2000; 2000US-0196460.

XX PR 11-APR-2000; 2000US-0196527.

XX PR 11-APR-2000; 2000US-0196528.

XX PR 11-APR-2000; 2000US-0547276.

XX PR 13-OCT-2000; 2000US-0240503.

XX PR 13-OCT-2000; 2000US-0240508.

XX PA (ARBO-) ARBOR VITA CORP.

XX PI Lu P, Garman JD, Candia AF;

XX DR WPI; 2001-375003/39.

XX PT P-PSDB; AAB99494.

XX PT Novel isolated cadherin-like asymmetry protein (CIASP) -3, useful for  
XX autoimmune disease -  
XX Disclosure; Fig 4A; 189pp; English.

CC The present invention describes an isolated polypeptide (I) comprising a  
CC nucleotide (nt) sequence (S1a) that has at least 90% identity to the  
CC sequence given in AAH41934 and is immunologically cross-reactive with  
CC the derived amino acid (aa) sequence (S1b) given in AAB99495 or shares a  
CC biological function with native CIASP-3. (II) has immunosuppressive,  
CC antiinflammatory, antirheumatic, antianemic, dermatological, uropathic,  
CC ophthalmological, antipneumatic, nephrotropic, antithyroid, antidiabetic,  
CC neuroprotective, antisthmatic and antibacterial activities, and can be  
CC used in antisense therapy, vaccine production and gene therapy. CIASP-3  
CC related sequences can be used in preventing or treating a CIASP-3  
CC mediated disease, preferably an autoimmune disease by inhibiting an  
CC immune response. The autoimmune disease is caused or exacerbated by  
CC increased activity of T helper 1 lymphocytes (TH1). AAH41893 to AAH41953  
CC and AAB99491 to AAB99507 represent sequences which are used in the  
CC exemplification of the present invention. CIASP-3 is localised in the  
CC chromosome location 1p31.1.  
XX  
SQ Sequence 4144 BP; 1252 A; 829 C; 890 G; 1173 T; 0 other;

Query Match 22.7%; Score 1446.6; DB 22; Length 4144;  
Best Local Similarity 64.5%; Pred. No. 6.8e-283;  
Matches 2239; Conservative 0; Mismatches 1199; Indels 33; Gaps 4;

```
Qy 2713 cgcattccgcccagcaactcttcaagagagctgtgctgtgagtggtgtcaagcagc 2772
Db 266 cgttaccacaaataaaagcttttccacagagagctgtgctgtgagtggtgttgcagt 325
Qy 2773 agtgcgtacgcagagcgcatcctccacagcgcgtgtctctctccagcagcatgtgaag 2832
Db 326 ggcagcgtctgggaatcagcttgcttgaacaaagcctggttctttttagtaaatgtgaag 385
Qy 2833 agtatgcgtctgacactgtctgtgcccagcagactagaacaccccgcaagctgtgcctc 2892
Db 386 agcatggtgcaccatttatacttaattatataactttagagctcccaagaaagcgtttt 445
Qy 2893 cccgagccttcctggagacatactgctctgtgtggtctgtggtgcgtggaggtatc 2952
Db 446 ccagaaagcttcaagagagacatgcagctctctgtcagcaagatgtctagtataatgt 505
Qy 2953 accgtgtccacaagagatgtgagctgtgcgcagacactcaagccagcctgtgtcttc 3012
Db 506 tcacgattcagaagagacagaagaattgttagagactcaataaagcctgtgactctt 565
Qy 3013 ctcaatgactgtgtgtctgtttagtgcagagagattgttttagcctataaagctctgc 625
Db 566 ctcaatgactgtgtgtctgtttagtgcagagagattgttttagcctataaagctctgc 625
Qy 3073 tacaagcaggtgtgcacgcgtctccagtcgtcccttaactccagcagcctgtgtgacctg 3132
Db 626 tataaacagtggtcttcaaaagcttactcattaccagatccagtgctgtgtgtcttg 685
Qy 3133 cgcattgaaatccacccgcatctgtgcagccagcagactcagtgaaaccttaacctccc 3192
Db 686 aggcgtgatttctcagatcatctgcagctcagatcagactatgtataaacttaacctcc 745
Qy 3193 tgcgtccctgtcaactcagactgcctccctccctctgtgtctcaccacctccag 3252
Db 746 tgcagcttacttactccacgcctcattccatcacaccttctgttcttctgtgaacatccag 805
Qy 3253 agctccacctctccagcagcccgagcccaagctgtaccagcagctgtgactagtg 3312
Db 806 agtctgatttctcagaaagtgtaacaaagaatgtgaataatgttgaattcc 865
Qy 3313 ggaccattccgagcagcaacttctcagctgtgtctctgtcgtgacagctgtgcactggcc 3372
Db 866 gtgccttcgcgcaacagcatatttgcagagactgtgttaacagagcgtgcgtcatc 925
Qy 3373 ctgaaactgaggtgtgaaggagctctgtgtgcaacaaagagccatcagtgctgtcac 3432
Db 926 ttgaacctgtgtcgaagagactgttgattgtcataaagaatcatcaatatgtacac 985
Qy 3433 agcctgctatgtgccaatgacatgaccccgctcagccgagcgcaactgtgaagcgctcgt 3492
```



Db 986 aattactctccagttcacagcaccagaccggtactcttgaccctccacagataaagctcga 1045  
Qy 3493 gtggcggagctgtacactgtccactgtcttcgattgcagcggtacacttgcacgctcat 3552  
Db 1046 gtggcagatgtgtatctactacccctgtattgtattatcaatggaactgttaccctcagcttat 1105  
Qy 3553 gaacttgcctgagggcccaagctcgaacggttcaagactgtgacctaaatgtgtactcagaaca 3612  
Db 1106 gattttacagaatacctcaacatcaacgagaagac-----caatttgtatagccactgat 1159  
Qy 3613 gaaggcgaagggtacattgcgggtataccatcaacccctctgtggtgcacatggtccattgtcgt 3672  
Db 1160 gattatgaagttgagagcggaaglatgataagaccagacggttgcacatggtcagatgcaggg 1219  
Qy 3673 ggcacctagaccctgtgtgtcccggtccagcatctcccaagggtccaccaacggtcttcgc 3732  
Db 1220 acatcgtgtccctcaactcaacaagaagcctgtgcagttctccctccacgttcaacgagtgagcag 1279  
Qy 3733 gcaagctgtgcccctctctgtctgagttcaagccggaacctgtgcgtgtgtgtgtgtgtgt 3792  
Db 1280 caacacactactacttcttcagcgaatacagttcgaagccttctgtactgttacttgggtc 1339  
Qy 3793 ctgaaaaaacaccgagccggtgctctctgtcagcgctgtggccacctgtgcacatcccccag 3852  
Db 1340 ctcaaaaatgcagatgaaacagcttctcaagaagctgtttacagatctctcagttctgtcag 1399  
Qy 3853 ctggtgagctgtctgtggaactgtgtacactgtgacctgtgcgtgtgccttctgtatcaaaaggtaa 3912  
Db 1400 ctaaacggtctattagatctgtcttactctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1459  
Qy 3913 aagccttctgaacgcatcaacaacagcctcaacatccaanaatctctgatatgaagcgcg 3972  
Db 1460 aaaaattgttgacaatgaatgaatagcttgaccttaagaataaagaatgagtgagcaag 1519  
Qy 3973 ctgagaggaagcattctgt 4032  
Db 1520 ctgtgaagagctatctctgt 1579  
Qy 4033 g-----aagaggagccggttctgt 4071  
Db 1580 ggaagctcgaagagagcccatctgtgaagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1639  
Qy 4072 cggaagagcgctcacacactgtgaaagcaaacctcagaacggtgtgtgtgtgtgtgtgtgtgtgtgt 4131  
Db 1640 aggaagaatatagtactctgt 1699  
Qy 4132 atggaacacgagcgctgt 4191  
Db 1700 atgtacaacgagacgtatgtatgt 1759  
Qy 4192 gacacactgtgagatcatcgt 4251  
Db 1760 gatcatltagagattgt 1819  
Qy 4252 ggggcaatgtcgtgaaggt 4311  
Db 1820 ggtgtgagtgctaaagtgctacttacaacatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1879  
Qy 4312 cagatgtgcctgtgcccacccagagggccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4371  
Db 1880 caaacacgttctgtctacaacagagagcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1939  
Qy 4372 gaagacacggt 4431  
Db 1940 gaagagacagagcaggt 1999  
Qy 4432 atcagcccatcgcagacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 4491  
Db 2000 atcgtgtatacaggtacacccacggt 2059  
Qy 4492 gagatcgccacacacttgcgt 4551

Db 2060 gagatttggataacttctgcaggtgttaaatgtcaagtgtacacatgtcatctctctgt 2119  
Qy 4552 gtgggagcagcagacacttcaatgtgaagcagacctgcagcttccactcaaacctccctc 4611  
Db 2120 gtgggcacatctcaaatlttaataagaatcttcaagcgttctctcaagaactatattgt 2179  
Qy 4612 acctatgtgagagacatctgt 4671  
Db 2180 acatatgttgaagaagatctgtgaattgt 2239  
Qy 4672 ctgtatttcaactctgacatgtatctcgcagacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4731  
Db 2240 ctgtatttcaatctcaatctgtatcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2299  
Qy 4732 gaccctgagatgtctacacccatgtacaaatgtgcccgggtgtccacaggggtctacagg 4791  
Db 2300 gatccctgaatgt 2359  
Qy 4792 gaccttcggtgtacactgt 4851  
Db 2360 gatctgt 2419  
Qy 4852 gctgt 4911  
Db 2420 gctgt 2479  
Qy 4912 ctggaagaccagcgcacactgtggt 4971  
Db 2480 ctgt 2539  
Qy 4972 gt 5031  
Db 2540 gttttgaagaatctgt 2599  
Qy 5032 tccgt 5091  
Db 2600 tctgt 2659  
Qy 5092 ttcacatgt 5151  
Db 2660 tttctgt 2719  
Qy 5152 ctgt 5211  
Db 2720 catgaagctaatctgt 2779  
Qy 5212 ttcacaaagatgt 5271  
Db 2780 ttcagcaaaatgt 2839  
Qy 5272 gt 5331  
Db 2840 gt 2899  
Qy 5332 ccactgtatcagaagctgt 5391  
Db 2900 cctgt 2959  
Qy 5392 tttgt 5451  
Db 2960 tttgt 3019  
Qy 5452 gactcaaaaaggtctacatccagatcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5511  
Db 3020 gatcttaacaaggtctacatccagatcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3079  
Qy 5512 ctcaagaccggt 5571  
Db 3080 atgaagagcagatcatctatttgcacaaaattacaaactctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3139  
Qy 5572 acgctgt 5631  
Db 3140 acaccttactttagatgt 3199





QY	3253	agctccacctcttcacgcgaagcccccgcgaaccccaagctgacagcatgttcgaactagtc	3312
Db	861	agctctgattctcttcacgaatgtatcaagaacccaagatgtcaataatgttgaattatcc	920
QY	3313	ggaacattcccgagacagacactcttcagatctggcctccctctgaagaaactggaactgccc	3372
Db	921	gtgcctcttcggccaacagcatattttggaagaaactgtgttaacagaagctggtctcat	980
QY	3373	ctcgaacactgaagcttgaaggggagatctctctgttgcaacaagaagggcatcagctgttcga	3432
Db	981	ttgaagccctgatgtcgaaggaactgtttgtgatgtgcataagaagaatgatcatatgtgaac	1040
QY	3433	agctctgtatgtgccaatgaacactgtgaccccccgtacgcgcagggacactgtgaagctcgt	3492
Db	1041	aatttactctccagatcaacagatcaagaccgcgcgttatctgtgacctccatagaataagctga	1100
QY	3493	gtgcgcgagatgtgaactcgcgcacactgcttcgatgtgacgaagatgaacttggccagctgcat	3552
Db	1101	gtggtccatgtgttatctactactgtgatgtgttatatcatgaaacactgtgacctcagctgat	1160
QY	3553	gaacttctgaaggggccaggttcagcgtgtcaaacactgtgacctatgtctgactcagaca	3612
Db	1161	gattttacagaagaactcaacatcaacaggaagaagac-----caattgtatagccaactgat	1214
QY	3613	gaagggcgaaggggaactcttcggtgtacatcaaacccctctgtgccaatgccaatgtcgtgt	3672
Db	1215	gattatgaaagtgtagaagcggaagtatgataaagccagaccggtgtgcattgtgcaatgcgaggg	1274
QY	3673	ggcccccctgaacccctctgctcccggtgcacacatctcccaaggggccacaaagcgcttcgc	3732
Db	1275	acatctgtccctctaactaatacaagggcctgtgcagtttccctcccaagtcacagatgtgcagg	1334
QY	3733	gcsagctgtgcctctctctctgtgagttcaagccggaaactgtgcgtgtgtgtctgtgagtg	3792
Db	1335	caacacacactcctcttcagcagaatcaagtcgaagccttttgatctgtctactcttggttc	1394
QY	3793	ctgtaaaaaacccgagccggcgcgtctctgtgcagcgtctggaactgtgactgtgacaactcccccag	3852
Db	1395	ctcaaaaaatgacagatgaaacagatctcccaagaaggtgtgttataagatctctccagcttctga	1454
QY	3853	cttgagacgtctgtgttgagactctgtgttaacctttgcgtgcttcgttcctgtgattacaagggaaa	3912
Db	1455	cttaacccgctatctagatctctgtctttatctctgtgtctctgtctctgtgagttaaaaagggaaa	1514
QY	3913	aagggccttgaacgcatacaacagcctccatcaatacaaaaatctctgatatgaatgaagcgcg	3972
Db	1515	aaagttcttgaagaagaaatgatatagctgtgacctttaagaatacaaaaagacatgagvgcaag	1574
QY	3973	cttgaggaagggccaattctgtgtgataccctgcagagctcgacaagaatgtgttcgycgaatcgtc	4032
Db	1575	ctctgaagaagcttaattctctgtggagcatalgtgcacgaagaatgtgatacggcgcgaagccga	1634
QY	4033	g-----aagggggccggtgttggaatccagaagaaagatgcgcgtg	4071
Db	1635	ggacacgctcgcagagaagagcccatcttgaaagtgtcctttggaagtcaagaataatttgaagttgg	1694
QY	4072	cggaaagagcgtcaacacacttgaaagcaaacctcaagaacgctgtgtaacaaagccaaagatgaa	4131
Db	1695	aggaagaagatatgactacactgcgtcaaaaacaagaagaagcttgacaataatcaagcgagag	1754
QY	4132	atgtaaacacagagggcctctgttgtaaggggaactcttgcaaacccgagcgaagcctagttgtctgt	4191
Db	1755	atctgaacaacgaagacacttgattgtatgaaacctctgtgttacaagaagaacaaactaatcatltta	1814
QY	4192	gaacacactggaagatcatcgtgcagacgggtgagtctttacagaagcccgggagagcgtcttg	4251
Db	1815	gataacttagaagatgtgtgttcacagcgcttctctgtaacggaatacccaagagagatctctc	1874
QY	4252	ggggcagctgtcgaaggtgtgtgtctgtacagcctctgggcagctgtgcgaagtgtccctctctgt	4311
Db	1875	ggctggagctgctaagaagtctactatacagaacgcatgctgcctgttaacaaagtgcagtttatcta	1934
QY	4312	caagcatggtcctgtgccaacccagaagggcctctgtgtccaagtcccggaagcgtgtgttcgag	4371

Db	1935	caaacctgttttgctacacagagagccttggtttcaagtttccgtacacttattgaa	1994
QY	4372	gagggacaaggagcgtgtgtgcgcgcctgtccgaggtctccacagacactgtggagccgc	4431
Db	1995	gaagagacagagcagtggtgtctgattttaaagctccacagctcttcgcacactgtgaagatag	2054
QY	4432	atcaacacacatccagacagccgcagcgctccgtctgatacgtctcatatgcagcaattc	4491
Db	2055	atcgtaacaataacgtttacaccccccagctgtccctcttaccctacctaagtaggcaaaattc	2114
QY	4492	gagatccgacacaactttgcccgtgtgtaagagatgcaggtcacacatgtctctctgcctctg	4551
Db	2115	gagatttggaaataacttttgcacaggtttaaataatgcaggtatccaatgtcacactctcctgt	2174
QY	4552	gtggggagcagcagccagaaacttcagttgaagcgccctcgagacgtttcacttcaaacacatcc	4611
Db	2175	gtgggcaacatccacgaatttttaataatgaagaattcttcaagacgttcttcaaaagactatgt	2234
QY	4612	aacctatcgtgagagagacatgagggcgtgcggagacagacacttcgcagagagcgtccagag	4671
Db	2235	acatatgtctgaagaagaatccgcgagatttgtaggaagaacaacattctctgtatccggtccg	2294
QY	4672	ctgatgttcaaacctgtcacatgtatcctgtacgagacacggtgtgaagatgaaagacacagag	4731
Db	2295	ctgggttttcaatcctccatagtattcttcttcgtatactgtgaataatgaagagacacagag	2354
QY	4732	gaacctgaagatgcctacacagccctcatgtacagaatttgcgcggggtccacaggtccacgg	4791
Db	2355	gacctctgaataatgttgatgtgcttcaaaagttaacgaatttgcacaagggtttccagacctcca	2414
QY	4792	gaacctgcgtacgtacgtgtgttcagacaatcagccggagacacgagacgtgggacacacac	4851
Db	2415	gactcgtgattgcacgtgtgttcagacaacatgagcagacgtcacactagaagaagaacatcat	2474
QY	4852	ggcgaagcgccccaagtgtgcatgtgtgcaagcgccgcctcgtgtgcatacctgccttg	4911
Db	2475	gcttaagactgcacagatgtctatgacaaactcagacagacactgtgtctgataatttggacatg	2534
QY	4912	ctcgaaagacacagcgccacacgtcccggtggggcttccgttcttccacagaaacatctacaac	4971
Db	2535	ctggaagagacccggaataatctctccgttggggatgtgttaaacattccagaactatcttacaat	2594
QY	4972	gtgtctaagagagtcgcgcacatctccgcagacacacccgttgcgcccgagcggagggctcttcg	5031
Db	2595	gttttaagaagatcctgcgttctccagatgtgtgtgtatctccagatgtgaagaagttatctgc	2654
QY	5032	tcgggggaagacattcaactgaactgtgggtgtgtgtatgtgtctgtgaaacagcgacgcgctac	5091
Db	2655	tctgtgaataacttactatgtatcagatgaactgtgtgtgtatctatgtgaacaagacgtcttcc	2714
QY	5092	ttcacacatggggggggtccttcaagacgggtgtgaatgaaggtcttcaagaacccctacccac	5151
Db	2715	ttctctatgtcgtgcactgtatgtatgaagcagtttaatgaagtttcaagaactattctctatt	2774
QY	5152	cttgaagacccacccgttactacacagaagactgtggccgggtgtcacggcacaactgcagagggc	5211
Db	2775	catgaagctaatccgggtatgcagaagaactatccacaattatcgaattgttaacaaggag	2834
QY	5212	ttcaccaagatctatgcacagaggttccgggtcgggtcggagcgggtgtctggagcgtattccgcg	5271
Db	2835	tctcgacaaaattgtcatatgaagtactgtcgtggagcggagatgttctgtgcacactatttctg	2894
QY	5272	gtgggcgtctcaagcggcccaacttgggtgaacctgtgataagaggaagtttgttaacaaggag	5331
Db	2895	gttgggtttttaatgtgaaccaggttccggggatttggatgaacaagaattgttttaacaaggag	2954
QY	5332	ccatcgtatccagagcgttgcagagatactccaacccgggtcgtgagaggttctacaacggagga	5391
Db	2955	ccgtgcaataacccaacttgcagaggtatctcaacagatttggagggattttaaaggagaaga	3014
QY	5392	tttggcagcagcgttctgtgagattatccaagaactctacccgttggacaggtccaagct	5451

```
Db 3015 ttggagagatgtgggtgaagatcaagaactctatcctgtgacaaagttaatta 3074
QY 5452 gactcaaaaagccctcaatccagatcagatgtgtgaaacgctactttgtacctagag 5511
Db 3075 gattccaagaagacatalttccagattacatgtgtgagccactactttgacacataagag 3134
QY 5512 ctcaagagacgggttgaccctacttgcgcgaactatgagcttgcacacttctgtctgc 5571
Db 3135 atgaagagacgaatccactatttcgcaaaaaattacaatcctgcgtcatgtatctgtcgt 3194
QY 5572 acgcgttcaacgcgcgagatggcgcgacacagggagctgcccgcgaacacacagctlaag 5631
Db 3195 acaccccttactttagatggtcggtgcacatggggaacttcatagaacaattcaaaagaag 3254
QY 5632 acgctgtcagccgcgcgaacgccttcctccatcacatcaagactgcgctgtgtgtccac 5691
Db 3255 accattctgactcgcgtcactccttctcttcatatataaacaagggttcaatgttcatcat 3314
QY 5692 cggagagagacgggtgtgtgacgcgagtggtggtgcacatgacagacatgacagaagaca 5751
Db 3315 aaagaagagacatctcttaacacaaatgaatgtgtcatgtgagacatgacagaaaaaagaca 3374
QY 5752 cggagagctgtgcttgcacacgcgaacccacacagatgtctaagatgtctacagatgtg 5811
Db 3375 caggagttgcatlgtgcaacacatcagatccgcgacaccccaaaatgcttcaagatgta 3434
QY 5812 ctccagagctctgttaagcccaacgcgacagaggtcccttgagagtggtgccagatgttt 5871
Db 3435 ctccaaggagatctgttaggcacacagatgaatcagggccttltggaagtlgtccagagttlt 3494
QY 5872 ttacagagatcccggaagaccccaagccttcgcgcgtcacacaaatgtgcgctctgc 5931
Db 3495 ctgtctgaataactactgtgtgacccaaagccttctagaacatcataataactgtgcactctgc 3554
QY 5932 ttaaggaactctgcagaagaattgtgagatgcgtgcggaaaaaataaggccctgtattgg 5991
Db 3555 tttaagatttctaataaagttgtgaagatgctctaagaaaaaataagagcttaattgg 3614
QY 5992 ccggacacagaagagatcacacgcgtgtgagcgagacgaactactgcgcgtctgcgagagct 6051
Db 3615 ccggttccaagaagagatcacaaaggaattg--gggaactatcttgcctttaaagagggcc 3672
QY 6052 ctgcagacccctgtcttaaccagcgcgctgcgccacgtgatagtgcaaccacccc---accgg 6107
Db 3673 ctacagcctcagatcacagaaagctcccaagtatcccaagccagatgtgtgttccctcg 3732
QY 6108 cctcagaactccttgacaagagcgaagttccgaaagcgacacactctgagc 6158
Db 3733 ccacagagatcctctcagtcgaatgagcttgcgaataatgagatctctaac 3783

RESULT 9
AAH41952
ID AAH41952 standard; DNA; 3173 BP.
XX
AC AAH41952;
XX
DT 30-AUG-2001 (first entry)
XX
DE CLASP-3 nucleotide fragment.
XX
KW Human; cadherin-like asymmetry protein; CLASP; CLASP-3; immune response;
KW cell surface molecule; transmembrane protein; immunosuppressive; vaccine;
KW antiinflammatory; antiarthritic; antianaemic; dermatological; utrophalic;
KW ophthalmological; antineumatic; nephrotropic; antidiabetic;
KW neuroprotective; antiautomatic; antibacterial; antisense therapy;
KW gene therapy; chromosome 1p31.1; ds.
XX
OS Unidentified.
XX
XX WO200142297-A2.
XX
XX 14-JUN-2001.
PD
```

```
XX 13-DEC-2000; 2000WO-US34171.
PF
XX 13-DEC-1999; 99US-0170453.
PR
PR 14-JAN-2000; 2000US-0176195.
PR 14-FEB-2000; 2000US-0182296.
PR 11-APR-2000; 2000US-0196267.
PR 11-APR-2000; 2000US-0196460.
PR 11-APR-2000; 2000US-0196527.
PR 11-APR-2000; 2000US-0196528.
PR 11-APR-2000; 2000US-0547276.
PR 13-OCT-2000; 2000US-0240503.
PR 13-OCT-2000; 2000US-0240508.
XX
XX (ARBO-) ARBOR VITA CORP.
XX
XX Lu P, Garman JD, Candia AF;
XX
XX WPI: 2001-375003/39.
XX
XX Novel isolated cadherin-like asymmetry protein (CLASP) -3, useful for
XX useful in preventing or treating a CLASP-3-mediated disease e.g.
XX autoimmune disease
XX
XX Disclosure: Fig 9A; 189pp; English.
XX
XX The present invention describes an isolated polypeptide (I) comprising a
XX nucleotide (nt) sequence (S1a) that has at least 90% identity to the
XX sequence given in AAH41934 and is immunologically cross-reactive with
XX the derived amino acid (aa) sequence (S1b) given in AAB99495 or shares a
XX biological function with native CLASP-3. (I) has immunosuppressive,
XX antiinflammatory, antiarthritic, antianaemic, dermatological, uropathic,
XX ophthalmological, antineumatic, nephrotropic, antihydroid, antidiabetic,
XX neuroprotective, antiautomatic and antibacterial activities, and can be
XX used in antisense therapy, vaccine production and gene therapy. CLASP-3
XX related sequences can be used in preventing or treating a CLASP-3
XX mediated disease, preferably an autoimmune disease by inhibiting an
XX immune response. The autoimmune disease is caused or exacerbated by
XX increased activity of T helper 1 lymphocytes (TH1). AAH41893 to AAH41953
XX and AAB99491 to AAB99507 represent sequences which are used in the
XX exemplification of the present invention. CLASP-3 is localised in the
XX chromosome location 1p31.1.
XX
SQ Sequence 3173 BP; 981 A; 624 C; 688 G; 880 T; 0 other;

Query Match 18.8%; Score 1195.8; DB 22; Length 3173;
Best Local Similarity 65.0%; Pred. No. 3.3e-232;
Matches 1846; Conservative 0; Mismatches 962; Indels 33; Gaps 4;

QY 3343 gggctcctgtcgtacggagctgacatggcctcgaaacctgaagctgaagggaattcctcg 3402
Db 4 ggaacttggtttcagaagctgctgctatttagacctgactgtcgaagactgttga 63
QY 3403 ttgcacaagaagagcatagcgtgctgacagagcctgtatgtgtgcacatgcacacgcc 3462
Db 64 ttgataagaagaatcatcaatataatgtgtaacaattactctcagctcagcactcagaccg 123
QY 3463 cgtacgcgcgagcgcaactgtgaagctcgtgtgacgcagctgtacacgcgcactgttcg 3522
Db 124 cgtactctgacccctcagataaagctcgcagtgatgcatatgtatctactccttgatggt 183
QY 3523 attgacaggaataccttgcacagcgtgcataacttgcgtgaagccagatcagagtgca 3582
Db 184 attatcatgaagaactggtaccctcagctgtatatttacaagaactcacacaatcaacaggga 243
QY 3583 agactgacctaatgctgtgactgcagacagaagaagcggaaggagcattggcggaacatc 3642
Db 244 agac-----caattgtatagccaatgatattatgaagtgagagcggaagatgata 297
QY 3643 aacctctgtgacatggccatgtgtgtgccccttagagcccttgctcccgggccagc 3702
Db 298 agccagaccgttgcacatgagcatcgcagggacatcgttccctcactaacaaggcctgvc 357
```

[illegible]

QY	4762	agattgtgccggggtcaccagggtccaccggactcttgctgacccctggttcagaacatg	4821
Db	1438	agaattgtccaaagggttataccagaccctccacagatccctgacatttgaccgtttgacgaacatg	1497
QY	4822	gccggagaagacacgcgaagcttgtaggcacaacacgcgcgaagcgccccaagtgcattggtgacacg	4881
Db	1498	gcagagcagaagcaccagaacgaagaacatcatgctgaagctgcacagtgctctagtctccatca	1557
QY	4882	gcgcgcctctgtgcttgagtaactctgcctctgctcgaaggccaagcccaactctgccgttggc	4941
Db	1558	gcagcactctgtctgaatttgacatctgtctgtaggagccggaataatctctctccca	1617
QY	4942	tgcgttccctccagaaacatccacccaacgcgtgtaagagatgcgcacatccgcacgac	5001
Db	1618	tgtgtacattccagatattcatcactaigtattttagaagaatctcggtctccagatgat	1677
QY	5002	atccctgtccgcgaagcagaagggtctctgctcgcgggaagcacttcaactctagctgggctg	5061
Db	1678	gtgtgattctccaatgatgaagaagatattcgcctcttgaaaaatacttacttgatccaagactt	1737
QY	5062	gttaggtgtgttgtaagaagacacgcggtacttaccatctggcgcggtctcagaagcggtg	5121
Db	1738	gttggtattactgtgaacaacagcagctgcttccctctctatgctgctgtatgtatgaagactt	1797
QY	5122	aattgagttcacagaacatccatcccaatccctgtaagcccaacgcgtgactacaagaagctg	5181
Db	1798	aattgaatttacaagaacttacttctccattcatgaagctatccggatgcagaagaacta	1857
QY	5182	gcgcggtgtgacagcgaacatgtgaaggagcctccacgaagtcatactgacagaagttccggc	5241
Db	1858	tcccaatcattcagtaataacttccaagaagcattccgcagaacaaattgttcataagattctggc	1917
QY	5242	tggagagcgcgtgttcggagcagatattccgcgttgagcttccacagcgcccaactctggtagc	5301
Db	1918	tggagagcagattgttgcaccatttctglttggttttatgtaaacaaagtctggggtat	1977
QY	5302	cttgatgtagcagaagattgtgtgacaaggagccatccagtcacgaagcttgccaagatctca	5361
Db	1978	tggagtgaaacaaatattgttcttaacaggagccgcgaataacccaactctgcagagataact	2037
QY	5362	caccgcgtctgagagagttcttccacacggaaggaatttgcgcagacgctgtgtgatttccaa	5421
Db	2038	cacagattggaaggaattttaaaggagaagatttggagagatgtgtgtgaagtaattccaa	2097
QY	5422	gactcttaccctgtgtgacacagtgccaagcttcttaccccaagaaggctcaatccaagatccg	5481
Db	2098	gactcctaatctctgtagacagtgatgaattgaatcttaacaaaggaatcaatccaagatcc	2157
QY	5482	tatgtgaaacggtacttatactccacagagctcacaaggaacgggttgaccacttttacaagc	5541
Db	2158	tatgtgtgagcctaactttgacacatagtagtgtaggaaggaacgaatcaactatttgacaaa	2217
QY	5542	aactatgttgcttcgacaattctctgttctgtcacgcgcttccacgcgcgaatgtggcgcaaac	5601
Db	2218	aattacaatcttcgtcagattcttctgactgtacgttacaaccttactttaagtgcgctgccat	2277
QY	5602	gggggagctgcgcggagacaacacaaagctgaagcgtctgtcagccgcgaacgccttccccc	5661
Db	2278	gggggaactcattgaacataacttcaaaaggagacacatctctgactacgtctcctacgtcttcc	2337
QY	5662	tacatcaagactgcacccgtgtgtgtgcacacgggaagagacggtgtctgacgcgaatggag	5721
Db	2338	tattatataaacaaggctcaatgtactcctcatagaagaagtgatctcttaaccacattgaa	2397
QY	5722	gttggtccatcgagacatgcagaaagaacacacggagcttggtccttgcacacgagcaggac	5781
Db	2398	gttgcattatgtagacatgtcagaaaaagacacaaaggaagttgttcatttcaacacatcaagat	2457
QY	5782	ccccccagatgctaaagtgtacaaatgtgtgtcttaagagcctcttaagggccacacgctggac	5841
Db	2458	ccccgcagaccccaaaatgtcttcagaatgtgtuactccaggatctgttgaagccacacagtgat	2517
QY	5842	cagggttccctctgaggtgtgccccagattgtttttagcagaagatccccggaagaccccaagctc	5901



Db 542 acaactactcgcaacccatgatctgctctccctctgcatcgtacgtacaaatgctgcatcagagcc 601  
Qy 3815 tcccgagcgctggtgccaacgagctgacactcccccgcggtggagagctctgttgacttgc 3874  
Db 602 tcatltaggaagtgatgttgcctgctgcacccaacgagcagctcaacagagattttgatctac 661  
Qy 3875 tgaaccttgctgctgctgctttagtacaaaggggaaaaagcgcttgatgaacgcatcaaca 3934  
Db 662 ttctcatctgtgtgtatgtttgttagataaggaacacgagttctcgacaaagtctagta 721  
Qy 3935 gctccacatccaacaaatcctctgatatgaaagcgcgcttagagagaaagccatctggtta 3994  
Db 722 ccccaagctctgcgaaggtcgaaggtgtcgaagcccgctggaagaggtcttgctgctgtg 781  
Qy 3995 ccatcgagctcgcaagaagatgttctgctgcaagtcgttgtagaga-----gcccgttgg 4048  
Db 782 gggaaggggccaagggggaatgtatgtgcgcggtgcctccagaaggaaacgcatctccag 841  
Qy 4049 ggaatccagaagaaagctgtgcctgtgcgaagaagctcacacactgtaagcaaacctcagacc 4108  
Db 842 gctcaaatgaaatttgagatggaagaagagcagacaacatttggtgcgcaagctaatgaga 901  
Qy 4109 ggcgtggcaagaagccaagagatgaaatggaacacgaagccttgtgtgaagggaaacctgaca 4168  
Db 902 agcttagataaaacaaagccgaagttagatcaagaagccttgatcagltggaactctgtcta 961  
Qy 4169 ccgaagcgaaagcctagctgtctctggaacacacactgtagagatcgtgcgaagcagctgtactt 4228  
Db 962 cagaagacactttaatcatctctgtgatatgtcaggaacaaatcatccagggcagactggtctc 1021  
Qy 4229 cagaagcccggaagagcgtctgtggggcaggtgctgtaaggttgtgtctgatacagcttgga 4288  
Db 1022 tggagctgtaaaagcagccctgtcggaaggtgtctggaaggtgtgctgtgaaattctctgaact 1081  
Qy 4289 gtgcgcaagagtgccctctctcttcacgacatgctgtgcaccccggaagggccctgtgtcca 4348  
Db 1082 gtatcatcgagatcacaccatcagctacatcagctcttgcaaacacccgtgctcatatgcaca 1141  
Qy 4349 agtcccgagagctgtctgtctggaagaggaacgagagctgtgtgcgaacgctgtgcttgagc 4408  
Db 1142 agtttgtagagcttactctctcgaaagagaggtggaacagtgcttcgacatctgttcaacaag 1201  
Qy 4409 tctctagaacactgtgtagcgcgcatacagaaccaaccgcgaagcagccgcctgcgtgt 4468  
Db 1202 tctctgaacacactgtcagacagagcatgtatgtatcccggaagcagccctgtgcacacccctt 1261  
Qy 4469 acctgtcatgtgcagaaacttgcagatcgtgcacaacttggccggtgtgaaagtatgcag 4528  
Db 1262 acctctcatgtaggttcaagtttctggagccaacgataatttgcgaagagtaaaagtatgcag 1321  
Qy 4529 tcaacatgtctctctctgctccctgtgtggagcgaacgagaaactcagttgaagagacactgc 4588  
Db 1322 taacacatgtccctgcatcttgtgtggagagcacaacgaactttaatgaaagagcacctgta 1381  
Qy 4589 gaagcttcaactaaacacatccctacatgtctgaagagagatgtgggtctgcggagacga 4648  
Db 1382 gaagatctctgtagacaatttgcctattcagaagaagacaaagcagcatcagatgacatc 1441  
Qy 4649 ccttgcgacagagcggtccagagacgtgtatcaacctgtacatgtacctgacgagacaagg 4708  
Db 1442 cttcttccaccacgggtgaggagaaactctctgtatctgtaatctgacatctatatatgacaag 1501  
Qy 4709 tgaaga tgaagagacacccggaagacccctgagatgctcatcgaacctcatgtacaaatgt 4768  
Db 1502 tgaagaatgtgggaatctcaggaagatccctggaatgtcttatgtatctcatgtacaaatgt 1561  
Qy 4769 cccggtgtactacaaggtcacaacgagactcggctgaacctgtgtgagaagaatcgtgcggga 4828  
Db 1562 ccaaggttaccaggtcatctctgcatctgctgagctgtgacacctgtctcagaagaatgtgagaga 1621  
Qy 4829 agcagcgcgagagctgtggcaacacacgacggagcccgccagttgcatgtgtcagcgcgccgc 4888

Db 1622 aacacacacaaagaaagtgtctacacggaaggtcgccatgtgctctgtgtgcacgcccgtgct 1681  
Qy 4889 tegtgtgtatgtaacctgcgcctgtctcgagagaccagcgccacactgcggtgggtgtgcgtt 4948  
Db 1682 tagtgtctgaagtatctgaacatgtcgtggaagacaaagcattacgtccgttggcagagtgtca 1741  
Qy 4949 ccttccagaacatctcatcaacggtctgtagaagagttccggcaactcttcgacgaacacccgt 5008  
Db 1742 gcttccagaatattcttcccaatgtgtcgtggaagagctgtgtgtctgtcgtggaacacccgt 1801  
Qy 5009 cgcccgacgaagagaggtctgtctccgggaagacacttcaactgtagcgtggtgtgtaggt 5068  
Db 1802 caactgacgaagatggtgtgtgtgcagggcagttacttcaacgaggtgctgtgttaggccc 1861  
Qy 5069 tgtgtgaacaggaacggtcgtacttccacatgtgggtggtggtccttagaagcggtgtgtagg 5128  
Db 1862 tccgtggaagagcgccggaagctcttcacgaacgggaggtcttatatgtagacagttataggg 1921  
Qy 5129 tctacaagaacctctcatcccatctcgtgaagcccaacgctgactatcaagaagctgtgcgg 5188  
Db 1922 tctacaagctgtgtcatctcccatctcagaagcgcatacgaagattccggaagctgacatca 1981  
Qy 5189 tgcacgcaaacctgtcagaagagccttcaccaaagatcatgacacagaagttccggtgtggagc 5248  
Db 1982 ctcaacagaagctgtcagaagagccttcgacagcatcgttaacaag-----gatcataga 2035  
Qy 5249 gctgtctcggaagctattccggtggtggtcttcaacggcccaacttcggtgtgactgtagt 5308  
Db 2036 gaatgtctggaacccaacttccgagctgtgttcttcttgatccaacatttgggatttggagt 2095  
Qy 5309 agcagagatttgtgtacaaagagccatcgtacacgaagcttgcagagaaattccacacagcg 5368  
Db 2096 aacagagatttgttcaaaagagcctgcaattacaaagcttcgcgagatctccacacatagac 2155  
Qy 5369 tggaggaattctcaacggaagagatttggcgacgagctgtgtgagatatacaaaagacctc 5428  
Db 2156 tagaggaacttbatgtatgaatgttgtgtgcaagatttgtggaagttgtatlaaagacccta 2215  
Qy 5429 accctgtgacaaggtccaaagcttgactacaaagagccctataccatacagtaagttgtg 5488  
Db 2216 ctccgtgtgacaaaacaaagctgtgataccttaacaaagccttaataacagatccacttgtgtg 2275  
Qy 5489 aacgctacttggatacctaagagctcaaggaacgggtgtaaccttatcttgcacgcaactatg 5548  
Db 2276 agccctacttggatgtagatgtagatgaaagcaggttcaacatacttggagaagaatttca 2335  
Qy 5549 ggtctgcgaacttctctgtctgcacgcccgttcaacgccggaatgtggcgacacagggagc 5608  
Db 2336 acctcggaggttcatgtacaccccgcgttcaaccccttggagggcggtcctcgggagagc 2395  
Qy 5609 tgcgcggaacaaacagagtaagagcgtctcagcacccgaaacagccttccctatactca 5668  
Db 2396 tgcattgagcagatacagaaggaacacagctctgcacacatbtgcagagccttccctacatca 2455  
Qy 5669 agactgcagctgtgtgtgcacacggggaaggaagagctgtgacgcgcaggtgtgaggtgtgca 5728  
Db 2456 agaccagagatcaggtcatccagaagaggaaggtgtgttttggacaacgattgaagtgtgca 2515  
Qy 5729 tgcagagacatgcagaagaagaacacgggaagctgtgccttgcaccagaggaagaccacag 5788  
Db 2516 ttgaagacatgaagaagaagacccctgcagtttagagttgcatatlaaacaggagcgcgct 2575  
Qy 5789 atgtcaagatgtacagaatgtgtcttgaaggtcgtctgtagggcccaacgggtgaaccaaaggtc 5848  
Db 2576 atgcagaagatgtctcaagaatgtgtctgcagaagcctctgtggagagctacgtgtaaatcaag 2635  
Qy 5849 cccctgggtgtggcccaagtttlttlaacagaagatcccggaagaccccaacgcttcccgcc 5908  
Db 2636 cactggaagtaggcccaagtggttttggcttgaatctcgtgtgttccaaatctcatcgcac 2695  
Qy 5909 atcacacaacaaatgtgcggtcgtctcaaggaactctgtcagaagaatgtgagatgtgcgtgc 5968  
Db 2696 atcacacaacagttgaggttatgtcttlaagaaattcatcagatgtgtgtgaaagcctgtag 2755





Db 738 atctgagcaatgctcgagagccacagctacctgccgttggcagctgacagcttcacagaata 797  
Qy 4961 tctcaccacagctgctagagagagtcgccatctcgcagacatctctgtcccgacagag 5020  
Db 798 ttcttcacaatgctcgagagagctgtgtctctgtagagcacccgttcaacttcgagag 857  
Qy 5021 agggcttcgctcgcgggaagcaacttcagctagcttgggctgtagagctgctggacaag 5080  
Db 858 atggaggtgtgcgacagagccagtaacttcacgagagatgctgtgtagcctcccgagacag 917  
Qy 5081 cagccggtactctcaccatctggcggtctctacagagcgggtgaatgaggtctacaagaacc 5140  
Db 918 ccgcgagagctcttcagcacgagggcttatatgagacagttaatgaggtctacagactg 977  
Qy 5141 tcattcccatctcggaagcccaacccgtgactacaagaagcttgcgcggtgcacgcaaac 5200  
Db 978 tcattcccatctcctagagagcgtcatcgagaattccggaagctgacacttcacacagaagc 1037  
Qy 5201 tgcagagagccttcaccagaatcatcaccagaggtcgcgtctggagcgcggttcgga 5260  
Db 1038 tgcagagagccttcgcagacagcatcgttaacaagatcata-----agagaatglttgaa 1091  
Qy 5261 cgtattccgcgttggccttctacgycgcgaacttcggtgacbtgga tgcagagagtttg 5320  
Db 1092 cctcattccgagttgttcttcttgatccaatttggggatgttgatgtagacagagatttg 1151  
Qy 5321 tgtacaagagagccatcgatcatcagaagctggcagaagattcttccacccgctggaggttct 5380  
Db 1152 tctcaaaagagccttgaatttaaccaagcttctctgaatctcaacttagacagagacattt 1211  
Qy 5381 acacgagagaatttggcgcagacagctgtctgagatataccaagaactcttcacgtctgagca 5440  
Db 1212 atggtcaatgttttggtagcagaatttggtagaagttagtaaaagctcactcctctgtgaca 1271  
Qy 5441 agtccaaagcttgactacaaaagcctacatccagatcagtatgttgaaacggtacttg 5500  
Db 1272 aaaccagaattgga tcttaaaagcctacatacagatcacttctgtgagacacttcttg 1331  
Qy 5501 atacctagagctcgaagagacggtgacacttctgacgcgaacttggcgctgcagact 5560  
Db 1332 atgagtagtagatgaaagaagaggtacacatacttctgagaagaattcaactcccgagagt 1391  
Qy 5561 tctgttctgacgcgcttcaacgcggtatggcgcgacacagggagagctgcccagacaac 5620  
Db 1392 tcatgtacacacaccgcttcaacccttgagggcggtcctcgggagagcgtgcatgagcagt 1451  
Qy 5621 acaagcgtaaagacgctgctcaagcacccagacacgcttccctacatccaagactgcgcatcc 5680  
Db 1452 acagaaggaacacagctctcgtacacatactgacgccttccctacatccaagacagatcat 1511  
Qy 5681 gttgtgaccccgaggaagacggtgctgacgcgagttgaggtgtggtcgcacgcgagacatgc 5740  
Db 1512 gctgtacccagaagagagagttgttcttgacacccgattgaagttgcatatgaagacatga 1571  
Qy 5741 agaagaagacacgagagctgtgcttgcacccgagcagagaccacacagatgtctaagatgc 5800  
Db 1572 agaagaagaccctgcagcttagcagttgacattaacacagagcgctgltgtcaagaatgc 1631  
Qy 5801 taagagtggtgtctcagaggtctgttagggcccccgtgaaacagaggttccccgtggaggttg 5860  
Db 1632 ttcgaagagtgctgcagaggtctgtggaagctactgtaatcagggacacacgtgaagttag 1691  
Qy 5861 cccaagtggttttagcagagatcccggaagaccacaaagctctcccgacacacaaacaaat 5920  
Db 1692 cccaagtggttttgcgtgaattccctgcgtgcatcacaactctatgcacatcaacaacagt 1751  
Qy 5921 tgcggtctgtctcaagactcttcagaagaatgtgagagatgcgtgcgcggaanaataaag 5980  
Db 1752 tggagttatgtcttaaggaattcatcatgagatgtgtgtgtagagtaaaataaacaagc 1811  
Qy 5981 cctgtatgtggccgagccagaagagtagacacgtgagctggggcgcaactactgcccgc 6040

Db 1812 gtctcatcagcgacacagaggaatatcagcaggaactccaagaataataaagc 1871  
Qy 6041 tgcggagagctctgaagccctgtcttaaccagcgccgcgcccagctg 6087  
Db 1872 taagaagaacctcagccaatgatcgagcggaataatccagaactg 1918  
  
RESULT 12  
AAC93501/c  
ID AAC93501 standard; CDNA: 1273 BP.  
XX  
AC AAC93501;  
XX  
AC 16-FEB-2001 (first entry)  
XX  
DE Human secreted protein gene 23 SEQ ID NO:33.  
XX  
KW Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;  
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
KW nootropic; neuroprotective; antibacterial; virocidic; fungicide;  
KW ophthalmological; vulnery; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorders; cancer; cardiovascular disorder;  
KW cardiac arrest; cerebrovascular disorder; nervous system disorder;  
KW Alzheimer's disease; ocular disorder; wound healing; skin aging; ss.  
XX  
XX Homo sapiens.  
XX  
XX MO200061626-A1.  
XX  
PD 19-OCT-2000.  
XX  
PF 06-APR-2000; 2000WO-US09066.  
XX  
PR 09-APR-1999; 99US-0128698.  
PR 20-JAN-2000; 2000US-0176926.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX (ROSE/) ROSEN C A.  
XX  
PI Rosen CA, Ruben SM, Komatsoulis G;  
XX  
XX WPI: 2000-619227/59.  
DR P-PSDB; AAB51849.  
XX  
XX New nucleic acid molecules encoding 49 human secreted proteins for  
PT diagnosing, preventing or ameliorating medical conditions and used for  
PT food additives or preservatives -  
XX  
XX Claim 1; Page 447; 516pp; English.  
XX  
XX Polynucleotide sequences AAC93479 - AAC93527 represent cDNA encoding  
CC human secreted proteins AAB51827 - AAB51875. Sequences AAB51876 -  
CC AAB51927 represent alternative polypeptides encoded by the genes, and  
CC amino acid sequences with which they share homology. The genes and  
CC proteins have activities dependent on the tissues and cells in which they  
CC are expressed. Examples of their activities include: immunosuppressive;  
CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;  
CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;  
CC virocidic; fungicide; ophthalmological; and vulnery. The secreted  
CC proteins, polynucleotides, antagonists and agonists may be useful in  
CC treating, preventing and/or diagnosing diseases and disorders such as  
CC autoimmune diseases e.g. Rheumatoid arthritis, hyperproliferative  
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders  
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia,  
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,  
CC infections caused by bacteria, viruses and fungi and ocular disorders  
CC e.g. corneal infection. The polypeptides can also be used to aid wound  
CC healing and epithelial cell proliferation, to prevent skin aging due to  
CC sunburn, to maintain organs before transplantation, for supporting cell  
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities, fat content, lipid, protein,  
CC carbohydrate, vitamins, minerals, cofactors and other nutritional



CC components. Oligonucleotides AAC93470 - AAC93478 and peptide AAB51826 are used in the isolation and characterisation of the proteins and CC polynucleotides of the invention.

xx Sequence 1273 BP; 254 A; 336 C; 366 G; 308 T; 9 other;

Query Match 13.0%; Score 829.4; DB 21; Length 1273;  
Best Local Similarity 87.3%; Pred. No. 3.3e-158;  
Matches 1005; Conservative 5; Mismatches 3; Indels 138; Gaps 4;

```
QY 5209 gcccttcacaaatcatatgacacagagttccggtctggagcggtgttctggagcgaatttc 5268
    |||||||
DB 1267 gcccttcacaaatcatatgacacagagttccggtctggagcggtgttctggagcgaatttc 1208
QY 5269 ccgcttgagcttctacagcgcccaactcggctgagcctgagatgagcaagagtttctacaa 5328
    |||||||
DB 1207 ccgcttgagcttctacagcgcccaactcggctgagcctgagatgagcaagagtttctacaa 1148
QY 5329 gagccatcatcagaagcttgacagagatctcacacggcctgagagagttctacacgag 5388
    |||||||
DB 1147 gagccatcatcagaagcttgacagagatctcacacggcctgagagagtttctacacgag 1088
QY 5389 agatttggagcagcagctgcttgagatttcaagaactctaccctgtgacaaagctccaa 5448
    |||||||
DB 1087 agatttggagcagcagctgcttgagatttcaagaactctaccctgtgacaaagctccaa 1028
QY 5449 ctgtactcacaagaagcctcacatccagatcacgtatgtgaacgcctacttatacctac 5508
    |||||||
DB 1027 ctgtactcacaagaagcctcacatccagatcacgtatgtgaacgcctacttatacctac 968
QY 5509 gagctaaagagcgggtgagctcaactttagccgaactaagggtctcgacatctctgttc 5568
    |||||||
DB 967 gagctaaagagcgggtgagctcaactttagccgaactaagggtctcgacatctctgttc 908
QY 5569 tgcacgccttcaacgcggatgagcgacacgggagcctcccgagcaacaacagcgt 5628
    |||||||
DB 907 tgcacgccttcaacgcggatgagcgacacgggagcctcccgagcaacaacagcgt 848
QY 5629 aagaagcttc-tcagacgcgaacacgccttccctcatcatcaagaactcgatcgtgtgtg 5687
    |||||||
DB 847 aagaagcttc-tcagacgcgaacacgccttccctcatcatcaagaactcgatcgtgtgtg 788
QY 5688 ccacgcggagaggaagctgtcagccagctgagagtggtgcacatcgaggacatgcagaaga 5747
    |||||||
DB 787 ccacgcggagaggaagctgtcagccagctgagagtggtgcacatcgaggacatgcagaaga 757
QY 5748 gacacggagagctgaccttgcacacgagacagaccacagatgctlaagatgtacagat 5807
    |||||||
DB 758 gacacggagagctgaccttgcacacgagacagaccacagatgctlaagatgtacagat 757
QY 5808 ggtgcttcaaggctctgtagggccacacgtgaacacaggttcccttgagggtggtccaggt 5867
    |||||||
DB 758 ggtgcttcaaggctctgtagggccacacgtgaacacaggttcccttgagggtggtccaggt 743
QY 5868 gtttttaagaagaatcccggaagaccccaagcgtcttcggagctcaacaacaatttggggt 5927
    |||||||
DB 742 gtttttaagaagaatcccggaagaccccaagcgtcttcggagctcaacaacaatttggggt 683
QY 5928 ctgtctcaaggaactctgcaagaatgtgaggaatgctgctgcggaataaagccctgtat 5987
    |||||||
DB 682 ctgtctcaaggaactctgcaagaatgtgaggaatgctgctgcggaataaagccctgtat 623
QY 5988 tgggcccggacagaagaagatgacacgctgagctgagcgaactactgcgcgtcggga 6047
    |||||||
DB 622 tgggcccggacagaagaagatgacacgctgagcgaactactgcgcgtcggga 563
QY 6048 ggtctgcaagcctcttcaacagcgcttgcacacgctgagatgagacccacacacgg 6107
    |||||||
DB 562 ggtctgcaagcctcttcaacagcgcttgcacacgctgagatgagacccacacacgg 503
QY 6108 cctcagaagaactccttgaacagagcaagtttccgaagagagacacttgcaccacaaga 6167
    |||||||
```

```
DB 502 CCTCAGGAACTCCTTGACAGAGCAAGTTTCCGAAGGCAAGACCTGTGAGCCCAACAAGA 443
QY 6168 ccaagctgtacacatagagagacacagacccgggacctgaactgtctgtgctgcgagggag 6227
    |||||||
DB 442 CCANAAGCTGTACCTAGAGAGAACAGACCCGGGGCTCAGCTGTCTGTGCTGCGAGGGAG 383
QY 6228 tctgcctgtgtgcccaactggtgctgtgggtgagacacacttacttgggctggccctct 6287
    |||||||
DB 382 TCTGCCCTGG-GYCCACTGGGCTGTGGGTGACACACTGTACTTNGGCT-GGCCCTCT 325
QY 6288 gcccctgtcccccatctgtgtgacatgacttccctcccttttaattaaatgatt 6347
    |||||||
DB 324 GCCCTGTGTGCCCATCTGTGTGTGACATGATCTTCCCTTTTAATTAAATGTT 265
QY 6348 ttataagca 6358
    |||||||
DB 264 TTTATAAGCAA 254

RESULT 13
AAC75563
ID AAC75563 standard; cdna; 586 BP.
XX
AC AAC75563;
XX
DT 08-FEB-2001 (first entry)
XX
XX Human ORF118 polynucleotide sequence SEQ ID NO:2235.
DE
XX Human; open reading frame; ORF; detection: cytostatic; hepatotropic;
KW antiparasitic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antihypertensive;
KW antiviral; antibacterial; antifungal; antineoplastic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200058473-A2.
PN
XX
XX 05-OCT-2000.
PD
XX
XX 31-MAR-2000; 2000MO-US08621.
PE
XX
XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CUBA-) CURAGEN CORP.
PA
XX
XX Shinketsu RA, Leach M;
PI
XX
XX WPI; 2000-602362/57.
DR
XX
XX P-PSDB; AAB41354.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
PS
XX
XX Claim 5; Page 1649; 5507pp; English.
CC
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORF open reading frames 1 to 3161. The ORF
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
```

CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antihaltic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;  
 CC antihydroid; and antianemic. The sequences can be used for determining  
 the presence of or predisposition to, or preventing or treating  
 pathological conditions associated with an ORF-associated disorder. The  
 CC nucleic acids can be used to express ORF proteins in gene therapy  
 vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease, to enhance  
 CC coagulation, to inhibit thrombosis, and as a contraceptive.  
 XX  
 SQ Sequence 586 BP; 122 A; 185 C; 156 G; 123 T; 0 other;

Query Match 7.4%; Score 468.6; DB 21; Length 586;  
 Best Local Similarity 98.0%; Pred. No. 2.5e-85;  
 Matches 465; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 519 gaatgactcccgagctgctcggtcccggaagacaccctcgaaagcagtgctc 578  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 88 gaatgactcccgagctgctcggtcccggaagacaccctcgaaagcagtgctc 147  
 QY 579 tagcatcttcgacatgaaagccttgagcagctactatctgtgcctctctagagcg 638  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 148 tagcatcttcgacatgaaagccttgagcagctactatctgtgcctctctagagcg 207  
 QY 639 ggcgcgcgcgaagatgctgacggcgcaatgaaccccttcgacggcagccgcccc 698  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 208 ggcgcgcgcgaagatgctgacggcgcaatgaaccccttcgacggcagccgcccc 267  
 QY 699 ggcctgctcaacccttacccggcagccttcgaagagatgaagcgtgaaagcgtctgagcgg 758  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 268 ggcctgctcaacccttacccggcagccttcgaagagatgaagcgtgaaagcgtctgagcgg 327  
 QY 759 ccgaagcgca-ccgcgcgcgacatttggaacaagatcttgtaagtgctctgcgtcca 817  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 328 ccgaagcgca-ccgcgcgcgacatttggaacaagatcttgtaagtgctctgcgtcca 387  
 QY 818 agtctgagatgaatctgacccatcttggaatcttggtctgtatgatgtgcggaga 877  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 388 agtctgagatgaatctgacccatcttggaatcttggtctgtatgatgtgcggaga 447  
 QY 878 aaaaagaatctcggaagacttctacttcgacctgaactcgactcatalgaagggcgctgc 937  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 448 aaaaagaatctcggaagacttctacttcgacctgaactcgactcatalgaagggcgctgc 507  
 QY 938 ttcggcctaigcagccccccttcgacatcttcacccctggccgcttcacatctctctg 997  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 508 ttcggcctaigcagccccccttcgacatcttcacccctggccgcttcacatctctctg 567  
 QY 998 tgaactacccctcaac 1012  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 568 tgactacccctcac 582

## RESULT 14

AA570846  
 ID AA570846 standard; cDNA; 981 BP.

XX AA570846;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #650.

KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX Homo sapiens.  
 OS WO200175067-A2.  
 PN WO200175067-A2.  
 XX 11-OCT-2001.  
 PD 30-MAR-2001; 2001WO-US08631.  
 PF 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 PA (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 P1 WPI: 2001-639362/73.  
 DR P-PSDB: ABG06659.  
 XX

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 PT  
 PS Claim 1; SEQ ID No 6650; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX

SQ Sequence 981 BP; 216 A; 274 C; 309 G; 182 T; 0 other;

Query Match 7.3%; Score 463.4; DB 23; Length 981;  
 Best Local Similarity 97.7%; Pred. No. 3.2e-84;  
 Matches 470; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1032 caagtggagaagtgcttaagcaagggacacgaatgagtgctgtgacctatagtt 1091  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 411 caagtggagaagtgcttaagcaagggacacgaatgagtgctgtgacctatagtt 470  
 QY 1092 gtgaaagaagtggacacagccaagaacaagaagctagagaagctgcgctggcggc 1151  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 471 gtgaaagaagtggacacagccaagaacaagaagctagagaagctgcgctggcggc 530  
 QY 1152 cgagcagttctgaccccgctggcgctaccgcagccttcctgctgtgaagcgctgca 1211  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 531 cgagcagttctgaccccgctggcgctaccgcagccttcctgctgtgaagcgctgca 590  
 QY 1212 ctggccaacatcgtgagcagcgctggcagcgtcggaacgtgactcgagggcga 1271  
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
 Db 591 ctggccaacatcgtgagcagcgctggcagcgtcggaacgtgactcgagggcga 650

QY 1272 ggcgcggcagacctgtgacagaccgcgcgtctgggggccccagagaccggcgagtaatgg 1331  
|||||  
Db 651 ggcgcggcagacctgtgacagaccgcgcgtctgggggccccagagaccggcgagtaatgg 710  
QY 1332 ggcgcgcgcctgcagcttcttctgtgtctccagccaaagcttaactgttacaacttct 1391  
|||||  
Db 711 ggcgcgcgcctgcagcttcttctgtgtctccagccaaagcttaactgttacaacttct 770  
QY 1392 taagcagagagctgagcagctcaatgacagaggaacctctcaagttctgtctgacatgag 1451  
|||||  
Db 771 taagcagagagctgagcagctcaatgacagaggaacctctcaagttctgtctgacatgag 830  
QY 1452 ggcgcgcgcctgcagcttctgtggcgagctacgtctgtgactgcccagctcaagatgacat 1511  
|||||  
Db 831 ggcgcgcgcctgcagcttctgtggcgagctacgtctgtgactgcccagctcaagatgacat 890  
QY 1512 t 1512  
Db 891 t 891

RESULT 15  
AAI61331  
ID AAI61331 standard; cDNA: 2427 BP.  
XX  
AC AAI61331:  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 5320.  
XX  
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Dreager Syndrome; chemotactic;  
KW chemoketic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukemia; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PE 26-DEC-2000; 2000MO-US34263.  
XX  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
PA (HYSE-) HYSO INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Dirmanac RT;  
XX  
DR WPI: 2001-442253/47.  
DR P-PSDB; AAM42175.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
PS Claim 1; SEQ ID NO 5320; 10078bp; English.  
XX  
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide

CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Dreager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 2427 BP; 716 A; 508 C; 524 G; 679 T; 0 other;

Query Match 7.0%; Score 448; DB 22; Length 2427;  
Best Local Similarity 67.0%; Pred. No. 5.2e-81;  
Matches 653; Conservative 0; Mismatches 315; Indels 6; Gaps 1;

QY 5114 aggcggtgaatgaggtctacagaacctcatccatccttgaagccaccgtgactaca 5173  
|||  
Db 1 agacagttaatgaggtctacagaacctgtgtcatcccatcctgaagcgcacgcgaattcc 60  
QY 5174 agaagctggccgcggtgcacgcgcaactgtcagagagccttcacaaagatacagaccaga 5233  
|||||  
Db 61 ggaagctgacactacactacacagcagcagagagccttcgaagatcgtaaacag- 119  
QY 5234 gtccgcgtggagcgcgtgttcggagctattccgcgtgtgcttacaagcgcaccac 5293  
|||  
Db 120 -----gatcataagagaatgtttgaacctacttcgcagtggttcttggatccaaat 174  
QY 5294 tcggtgacctgagatgagcagagattgtgtacaagagacatcgatacagaactggcag 5353  
|||  
Db 175 ttggagatttgatgaaacaggaattgttctacaagagccctgcacattacaagcttctg 234  
QY 5354 agatctcacaccgctgtgagagattctacacgagagatgtggagcagctgtgtaga 5413  
|||||  
Db 235 agatctcacacagactagagcattttatgttcattgttggcagaaatttgggaag 294  
QY 5414 ttatcaagactcttaccctgttgacaagtlccaaagcttgactccaagaagcctacatcc 5473  
|||||  
Db 295 tgattaaagactccactcctgttgacaacaaacgaattgattacataacagcctacatcc 354  
QY 5474 agatcagttatgtgaaacgcttctgtatccatcagagctaaagacccggttgacctact 5533  
|||||  
Db 355 agatcactttgtgagccctacttgaatgatatgaatgaagaagaggttcacatact 414  
QY 5534 ttgacgcgaactatggtctgcacattcctgttctgcagccgttcaacgcgcatgggc 5593  
|||||  
Db 415 ttgagaagaattcaactcctcggagaggttcattgatacacaccacccgttcaaccttgaggggc 474  
QY 5594 gcgcacacggggagctgcccagacacaacagcgtlaaagcgtgctcgaacgcagcaag 5653  
|||||  
Db 475 ggccctcggggagagctgcatgacagatagcagtaagaagaacacagctcctgacata 534  
QY 5654 ccttcctctacataagactcgcacatcgctgtgtgcccacccgggaagagacggtgtgagc 5713  
|||||  
Db 535 ccttcctctacataagacccaagcagcagcgttcacagaagagagaggtgttttgacac 594  
QY 5714 cagtgtgagtgccatcgcagagacatgacagaaagacacggagacgtgcttgcacgcg 5773  
|||||  
Db 595 cgattgaagttgaccttgaagaacatgaagaagaagacccttcagttgacgttgcat 654  
QY 5774 agcaggaaccacagatgataagatgtctacagatgtgtcctcaaggtctgttagagccca 5833  
|||||  
Db 655 accaggaagcgcgtgatgacaagatgtctcagatgtgtctgcaaggtctgttgagacta 714  
QY 5834 ccgtgaaccagagttcccttgagagtggtggccacagtggttttgcagagatcccggaagacc 5893  
|||||  
Db 715 ctgttaactcagagaccactggaagtagcccaagtggttttgcgtgaattcctgcgtgac 774  
QY 5894 ccaagctcttcggatcacacaacaattgctgctgtctcaagagacttcgcagaagaat 5953

```

Db 775 caaaactctatcgacatcacacaagttgaggtatgctttaaggaattcatcatgagat 834
QY 5954 gtgaggtgctgctgcggaataaagccctgtgttgccgcgacccaagaagatcacc 6013
Db 835 gtggtgaagctgtagaagaaacaagcgtctcatcacgacgacccaagaggaataatcacgc 894
QY 6014 gtgagctgagcgcaactactgcccctgcygagagctctgcagccccctgcttaccagc 6073
Db 895 aggaactcaaaaagaactataacacagctaaagagaacctcagccaatgatcgagcgga 954
QY 6074 gcctgcccagctg 6087
Db 955 aaatccagaaactg 968

```

Search completed: October 5, 2002, 08:20:24  
 Job time: 20081 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 02:30:18 ; Search time 156.01 Seconds  
(without alignments)  
10032.541 Million cell updates/sec

Title: US-09-736-968A-1  
Perfect score: 6372  
Sequence: 1 gacgcgagacattgctgc.....aagcaaaaaaaaaaaaaa 6372

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTCUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	67	1.1	1931	2 US-09-130-114-2	Sequence 2, Appl
2	65.6	1.0	2580	4 US-09-050-863-2	Sequence 2, Appl
3	65.6	1.0	2580	4 US-09-359-081-2	Sequence 2, Appl
4	65.6	1.0	5452	2 US-09-130-114-1	Sequence 1, Appl
5	65.6	1.0	9600	4 US-08-910-647-1	Sequence 1, Appl
6	65.6	1.0	10596	1 US-07-884-811-15	Sequence 15, Appl
7	65.6	1.0	10596	1 US-07-885-971-15	Sequence 15, Appl
8	65.6	1.0	10596	1 US-08-087-783A-15	Sequence 15, Appl
9	65.6	1.0	10596	1 US-08-194-088B-15	Sequence 15, Appl
10	65.6	1.0	10596	2 US-08-194-087-15	Sequence 15, Appl
11	65.6	1.0	10596	5 PCT-US93-04646-15	Sequence 15, Appl
12	61.4	1.0	7218	1 US-08-232-463-14	Sequence 14, Appl
13	56.2	0.9	1908	3 US-09-046-992-1	Sequence 1, Appl
14	52.8	0.8	7218	1 US-08-332-463-14	Sequence 14, Appl
15	51.2	0.8	1869	2 US-08-356-86-15	Sequence 15, Appl
16	51.2	0.8	1931	2 US-09-130-114-2	Sequence 15, Appl
17	50.8	0.8	2493	1 US-07-977-434-5	Sequence 5, Appl
18	50.8	0.8	2493	1 US-08-458-819-5	Sequence 5, Appl
19	50.8	0.8	2493	5 PCT-US91-07035-5	Sequence 5, Appl
20	49.6	0.8	1191	3 US-09-046-992-3	Sequence 3, Appl
21	49.6	0.8	1389	1 US-08-484-438-41	Sequence 41, Appl
22	49.6	0.8	1797	1 US-08-463-163-2	Sequence 2, Appl
23	49.6	0.8	2760	2 US-08-743-637B-20	Sequence 20, Appl
24	49.6	0.8	2760	3 US-08-526-840B-20	Sequence 20, Appl
25	49.6	0.8	2760	3 US-08-748-170A-3	Sequence 1, Appl
26	49.6	0.8	2760	3 US-09-047-148-1	Sequence 1, Appl
27	49.6	0.8	5727	5 PCT-US91-02954-13	Sequence 13, Appl

28	48.4	0.8	3527	2 US-08-909-965C-7	Sequence 7, Appl
29	47.6	0.7	489	4 US-09-199-637A-430	Sequence 430, App
30	47.6	0.7	951	4 US-09-199-637A-432	Sequence 432, App
31	47.6	0.7	1050	4 US-09-199-637A-428	Sequence 428, App
32	47.6	0.7	1780	2 US-08-933-821-5	Sequence 5, Appl
33	47.6	0.7	1780	3 US-08-960-507-5	Sequence 5, Appl
34	47.6	0.7	1780	4 US-09-136-828-5	Sequence 5, Appl
35	47.6	0.7	2012	1 US-08-235-838-13	Sequence 13, Appl
36	47.6	0.7	2012	1 US-08-235-838-15	Sequence 15, Appl
37	47.6	0.7	2012	2 US-08-465-473B-13	Sequence 13, Appl
38	47.6	0.7	2012	2 US-08-465-473B-15	Sequence 15, Appl
39	47.6	0.7	42235	4 US-09-199-637A-1	Sequence 1, Appl
40	47.2	0.7	1268	2 US-08-510-646B-2	Sequence 2, Appl
41	47.2	0.7	1268	3 US-08-510-646B-2	Sequence 2, Appl
42	47.2	0.7	1268	4 US-09-231-818-2	Sequence 2, Appl
43	47.2	0.7	5392	2 US-08-403-852D-1	Sequence 1, Appl
44	47.2	0.7	5392	3 US-08-510-646B-1	Sequence 1, Appl
45	47.2	0.7	5392	4 US-09-231-818-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-09-130-114-2  
; Sequence 2, Application US/09130114  
; Patent No. 5976807  
; GENERAL INFORMATION:  
; APPLICANT: Horlick, Robert A.  
; APPLICANT: Robbins, Alan B.  
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes  
; TITLE OF INVENTION: From Multiple Transfected Epilomes  
; FILE REFERENCE: 0867/ID903US1  
; CURRENT APPLICATION NUMBER: US/09/130,114  
; CURRENT FILING DATE: 1998-08-06  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1931  
; TYPE: DNA  
; ORGANISM: EBNA  
US-09-130-114-2

Query Match	1.1%	Score 67;	DB 2;	Length 1931;
Best Local Similarity	45.2%	Pred. No. 1.2e+05;		
Matches 286;	Conservative 0;	Mismatches 345;	Indels 2;	Gaps 1;
QY 2652	catcagcagcagcaacccctgactcgcgtgagccctgctctgagtcgaggttc	2711		
DB 325	cgctccgctccctccctccctccctccctccctccctccctccctccctccctccct	384		
QY 2712	ccgcactcctgagcagcaagctgctcaagagagctgctcagtgagtgagtcagag	2771		
DB 385	cccgctccctccctccctccctccctccctccctccctccctccctccctccctccct	444		
QY 2772	ccagtcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	2831		
DB 445	cccgctccctccctccctccctccctccctccctccctccctccctccctccctccct	504		
QY 2832	gagtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	2891		
DB 505	cgctccgctccctccctccctccctccctccctccctccctccctccctccctccct	564		
QY 2892	ccccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	2951		
DB 565	ccccgctccctccctccctccctccctccctccctccctccctccctccctccctccct	624		
QY 2952	caaccgctcccaagagatgagagctgagcagcagcagcagcagcagcagcagc	3011		
DB 625	ccccgctccctccctccctccctccctccctccctccctccctccctccctccctccct	684		

[illegible]

```

1      RESULT      2
2      US-09-050-863-2/c
3      Sequence 2, Application US/09050863
4      Patent No. 6114111
5      GENERAL INFORMATION:
6      APPLICANT: Lao, Ying
7      APPLICANT: Hiang, Betty
8      APPLICANT: Payan, Don
9      TITLE OF INVENTION: Mammalian Protein Interaction Cloning
10     TITLE OF INVENTION: System
11     NUMBER OF SEQUENCES: 5
12     CORRESPONDENCE ADDRESS:
13     ADDRESSEE: Flierl, Hobbach, Test, Albritton & Herbert
14     STREET: 4 Embarracadero Center, Suite 3400
15     CITY: San Francisco
16     STATE: CA
17     COUNTRY: USA
18     ZIP: 94111-4187
19     COMPUTER READABLE FORM:
20     MEDIUM TYPE: Floppy disk
21     COMPUTER: IBM PC compatible
22     OPERATING SYSTEM: PC-DOS/MS-DOS
23     SOFTWARE: PatentIn Release #1.0, Version #1.30
24     CURRENT APPLICATION DATA:
25     APPLICATION NUMBER: US/09/050.863
26     FILING DATE: 30-MAR-1998
27     CLASSIFICATION:
28     ATTORNEY/AGENT INFORMATION:
29     NAME: Silva, Robin M.
30     REGISTRATION NUMBER: 38,304
31     REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
32     TELECOMMUNICATION INFORMATION:
33     TELEPHONE: (415) 781-1989
34     TELEFAX: (415) 943-8711
35     INFORMATION FOR SEQ ID NO: 2:
36     SEQUENCE CHARACTERISTICS:
37     LENGTH: 2580 base pairs
38     TYPE: nucleic acid
39     STRANDEDNESS: unknown
40     TOPOLOGY: unknown
41     MOLECULE TYPE: DNA
42     US-09-050-863-2

```

Query Match	1.0%;	Score 65.6;	DB 3;	Length 2580;
Best Local Similarity	47.0%;	Pred. No. 2.7e-05;		
Matches 235; Conservative	0;	Mismatches 264;	Indels 1;	Gaps 1

Dy 2778 cgtacgagaggccatcctccagcagcgcttgttcttcttcacgctaagtgaagatat 2837  
| | | | | | | | | | | | | | | |  
Db 1381 cctccgcgaccggcgctccacactcctgctcctgcccccctcctgctcctgccctctcmct 1322

QY	2838	ggagtgagcaatgctgtgtgtggcaggagcatagacacacccggaagtgtggttcccg	2897
Db	1321	gctcttgccctctctgcccctctctgcttctgcccctctgcccctctctctctctct	1262
QY	2898	acgcttccttgagacacatacatctgcttgtgtgtgtctgtgtggccttgagatcataccg	2957
Db	1261	cctggccctctctgctctgcccctctgcccctctgcccctctgcccctctgcccct	1207
QY	2958	tgtccacaagaatgtgtgagctgtgcgcgaagcaactcaagccagacttgccttcctacg	3017
Db	1201	ctctctgctctgcccctctctgcccctctgcccctctgcccctctgcccctctctct	1142
QY	3018	tgaactctgtcctgtgtgtgacggaggttgtcttcagcctgtgtccggggccactaca	3077
Db	1141	gcccctctgcccctctctgctctgcccctctgcttctgcccctctgcttctgcccct	1082
QY	3078	gcaggtgtgacagcagcgtctccagtcgtcccttaaccagacagcctgtgacctgtgcgt	3137
Db	1081	ctctctctctgcccctctctgcccctctgcccctctgcccctctgcccctctgcccct	1022
QY	3138	ggaattcaaccgcatctctgtgcagacagacatacgtgacctctcaactcctctgct	3196
Db	1021	gcccctctgcccctctctgcccctctgcccctctgcccctctgcccctctgcccct	962
QY	3197	ggcccctgtgcaactctcaagcctctgccttccctctgtgtctctcaacacacttcccaagct	3256
Db	961	gcccctctgcccctctctgcccctctgcccctctgcccctctgcccctctgcccct	902
QY	3257	gcacctcttcagacaaagcc 3276	
Db	901	ctctctgctctgcccctctcc 882	

```

RESULT 3
US-09-359-081-2/c
; Sequence 2, Application US/09359081
; Patent No. 6316223
; GENERAL INFORMATION:
; APPLICANT: Lao, Ying
; Hiang, Betty
; Payan, Don
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning
; System
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehm, Hohbach, Test, Albritton & Herbertt
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/359,081
; FILING DATE: 22-Jul-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/050,863
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2580 base pairs

```

TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-359-081-2

Query Match 1.0%; Score 65.6; DB 4; Length 2580;  
Best Local Similarity 47.0%; Pred. No. 2.7e-05;  
Matches 235; Conservative 0; Mismatches 264; Indels 1; Gaps 1;

```
QY 2778 cgtacgagagagcattccacagacgctgtgtctcttcacagctatgtgtgaagat 2837
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1381 CCGTCTGAGACCCGCGCTCACCCTCTGCTCCCTGCGCCCTCGTCTCCCTCCTCCT 1322
QY 2838 ggcgtgacactgtctgtgacgagactagacaccccgcaagcttcgctcccg 2897
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1321 GCTCTGCCCCCTCTGCCCCCTCTGCTGCCCCCTGCTGCCCCCTCTGCTGCCCCCT 1262
QY 2898 acgcttcctgagacatcactgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2957
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1261 CCGGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1202
QY 2958 tgtccaaagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3017
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1201 CCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1142
QY 3018 tgaacctgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3077
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1141 GCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1082
QY 3078 gcaagtgtgacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3137
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1081 CCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1022
QY 3138 ggaattacaccgcatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3196
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1021 GCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 962
QY 3197 gccctctgtacatcctcagctcctccctcctcctcctcctcctcctcctcctcct 3256
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 961 GCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 902
QY 3257 ccacctctccagcagcc 3276
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 901 CCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 882
```

## RESULT 4

US-09-130-114-1  
; Sequence 1, Application US/09130114  
; Patent No. 5976807  
; GENERAL INFORMATION:  
; APPLICANT: Horlick, Robert A.  
; APPLICANT: Damaj, Bassam B.  
; APPLICANT: Robbins, Alan K.  
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes  
; FILE REFERENCE: 0867/ID903051  
; CURRENT APPLICATION NUMBER: US/09/130,114  
; CURRENT FILING DATE: 1998-08-06  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 5452  
; TYPE: DNA  
; ORGANISM: VERBA  
US-09-130-114-1

Query Match 1.0%; Score 65.6; DB 2; Length 5452;  
Best Local Similarity 47.0%; Pred. No. 3.5e-05;

Matches 235; Conservative 0; Mismatches 264; Indels 1; Gaps 1;

```
QY 2778 cgtacgagagagcattccacagacgctgtgtctcttcacagctatgtgtgaagat 2837
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1424 cctctgaccccgagctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1483
QY 2838 ggcgtgacactgtctgtgacgagactagacaccccgcaagcttcgctcccg 2897
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1484 gctctgcccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1543
QY 2898 acgcttcctgagacatcactgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2957
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1544 cctgcccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1603
QY 2958 tgtccaaagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3017
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1604 cctctgtctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1663
QY 3018 tgaacctgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3077
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1664 gccctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1723
QY 3078 gcaagtgtgacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3137
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1724 cctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1783
QY 3138 ggaattacaccgcatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3196
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1784 gccctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1843
QY 3197 gccctctgtacatcctcagctcctccctcctcctcctcctcctcctcctcctcct 3256
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1844 gccctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1903
QY 3257 ccacctctccagcagcc 3276
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1904 cctctgtctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1923
```

## RESULT 5

US-08-910-647-1/c  
; Sequence 1, Application US/08910647  
; Patent No. 6251433  
; GENERAL INFORMATION:  
; APPLICANT: Zuckermann et al.  
; TITLE OF INVENTION: Compositions and Methods for  
; TITLE OF INVENTION: Polynucleotide Delivery  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 94608-2916  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/910,647  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fujita, Sharon M.  
; REGISTRATION NUMBER: 38,459  
; REFERENCE/DOCKET NUMBER: 1218, 002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 923-2706  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 9600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
US-08-910-647-1

```

Query Match	1.0%	Score 65.6;	DB 4;	Length 9600;
Best Local Similarity	47.0%;	Pred. No. 4.3e-05;		
Matches 235; Conservative	0;	Mismatches 264;	Indels 1;	Gaps 1;

QY	2778	cgagcgagagacatccttcacgaagcagccggtgtctcttcacagctcaygtgaagat	2837
Db	1427	ccctctgcagaccggggctccacactccgcgtccgcccctccgctcccccctccct	1368
QY	2838	ggagctgcacctgtctgttcggcagcagactagacacaccgcgaagtgcgtcccg	2897
Db	1367	gcctctggccctctctgcgcccctccggtccggcccctcctgcccctcctcctccct	1308
QY	2898	acgcttccttgaagacatcaactcgtccttgtgtggctctgtggacctgaagtcacccg	2957
Db	1307	ccgcgccctctcgcgtccctgcgcccctccctccgcccctccgcgtccgcccctccctccct	1248
QY	2958	ttgcacaaagatgttgaagcttgcgcagacacacacgcagccagcgtcttccttcag	3017
Db	1247	ccctcctgcgtcctgcgcccctcctgcgcccctccgctccgctccgcccctcctcctcct	1188
QY	3018	tgacctctctgccttgcgttgaacggggcttgccttcacgctgtgtccgggccactaca	3077
Db	1187	gccctctcctgcgcccctccgtccctgcgcccctccgctccgcccctccgctccctccct	1128
QY	3078	gcaggttgcacagcggtctccacgtcgtccccaatccacagacgcctgcgtacccctgcgat	3137
Db	1127	ccctcctcctgcgcccctcctgcgcccctccgctccgcccctccgctccgcccctcctcct	1068
QY	3138	ggaattcaaccggatcctctgtgcagcagagacgaactaagtgcacctcaact-ccccgtct	3196
Db	1067	gccctctcctgcgcccctccgtccctgcgcccctccgctccgcccctccgctccctccct	1008
QY	3197	ggcccccgtcacatccacagcctgcgcccctcccccctgtgtcctccacacacccacagagct	3256
Db	1007	ggccctcctgcgcccctcctcctgcgtccgtccgcccctccgcccctccgctccgcccct	948
QY	3257	ccaactctccagcgaagcc 3276	
Db	947	ccctcctcctcctgcgcccctcc 928	

RESULT 6  
 US-07-884-811-15/c  
 Sequence 15, Application US/07884811  
 Patent No. 5316921  
 GENERAL INFORMATION:  
 APPLICANT: Godowski, Paul J., Lohker, Nathalie A. Mark, Melanie R  
 TITLE OF INVENTION: SINGLE-CHAIN HEPATOCTE GROWTH FACTOR VARIANTS  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Genentech, Inc.  
 STREET: 460 Point San Bruno Blvd  
 CITY: South San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94080  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: path (Genentech)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/884,811  
 FILING DATE: 19920518  
 CLASSIFICATION: 530

```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
:
: FILING DATE:
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Dregier, Ginger R.
: REGISTRATION NUMBER: 33,055
: REFERENCE/DOCKET NUMBER: 755.1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/225-3316
: TELEFAX: 415/952-9881
: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10596 bases
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-07-884-811-15

```

Query Match	1.08;	Score 65.6;	DB 1;	Length 10596;
Best Local Similarity	47.08;	Pred. No. 4.4e-05;		
Matches 235; Conservative	0;	Mismatches 264;	Indels 1;	Gaps 1;

QY	2778	cgatcgcgagggacatcctctccagaagcgcgtgltctctctccagctcaatggtgaagatcat	2837
Db	2961	CTCTCTGACACCGGCGCTCCACCTCTCTGCTCTGCGCCCTCTGCTCTGCGCCCTCTCT	2902
QY	2838	ggcgcgtgacactgctgctctgtgcccagcagctatagacacaccgcgaaactgtgcgtcccg	2897
Db	2901	GCTCTCTGCGCCCTCTGCTCTGCGCCCTCTGCTCTGCGCCCTCTGCTCTGCGCCCT	2842
QY	2898	acgcttcctctggaagaaactcaactcgtccttgtgtggtcctctgtggcctcggaggtataccg	2957
Db	2841	CTTGCGCCCTCTCTGCTCTGCGCCCTCTGCGCCCTCTCTGCTCTGCGCCCTCTGCGCCCT	2782
QY	2958	tgtccacaagaatgtgtgagctctgacgcagcagcactcaagcagcagctgtctctccctacg	3017
Db	2781	CTCTCTGCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCTCT	2722
QY	3018	tgaactctgttcctctgtgtgacacggggcttgtgtctctagcctcgtgtccgggcgccaataaa	3077
Db	2721	GCCCTCTCTCTGCGCCCTCTGCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCT	2662
QY	3078	gcaggtgtgcacagcggcgtccacgtcgtccctaatccagaagccctcgtgacccctgcgat	3137
Db	2661	CTGCTCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCTCT	2602
QY	3138	ggaaattacccgcatacctctgtgcagccagagcaactaagltgacctcaact-cccctgct	3196
Db	2601	GCCCTCTCTGCGCCCTCTGCGCCCTCTGCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTCT	2542
QY	3197	ggccctcgttaactcaagcctcgtgcctcccccctcgtgtccttcacacaaactccagagct	3256
Db	2541	GCCCTCTCTGCGCCCTCTGCTCTGCTCTGCGCCCTCTGCGCCCTCTGCTCTGCTCTGCGCCCT	2482
QY	3257	ccacctctccagccaagcc 3276	
Db	2481	CTCTCTGCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCTCT	2462

```

RESULT 7
US-07-885-971-15/c
; Sequence 15, Application US/07885971
; Patent No. 5328837
;
; GENERAL INFORMATION:
;
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCTYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
;
; NUMBER OF SEQUENCES: 21
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco

```





```

Db      2661 CCTGCTCCTGCCCCCTCGCCGCCCTTCGTACCCCCTCCTCCTGCTCCTAGCCCCCTTCGTACTT 2602
Qy      3138 ggaattcacccgcatctcgtgtagagccacgagactaaatgaacctcaact-ccccgtct 3196
Db      2601 GCCCTCTCTGCACCCCTCGGCCCTTCCTGCTCCTTGCCCTCCTCGTCACTCGCCCTTCCT 2542
Qy      3197 gcccccgttaacctaacgcctcgccctccctcctcgtgttcttcaccacacccccaagtct 3256
Db      2541 GCCCTCTCTGCACCCCTCGTCTGCTCTGCTGTGCCCCCTCTGCCCCCTCTCTGCTCTGCCCCCT 2482
Qy      3257 ccacctctctccagccaagcc 3276
Db      2481 CCTCTGCTCCTCGGCCCTTC 2462

RESULT          9
US-08-194-088B-15/c
; Sequence 15, Application US/08194088B
; Patent No. 5580963
; GENERAL INFORMATION:
; APPLICANT: Godowsky, Paul J., Lokker, Nathalie A., Mark, Melanie R.,
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,088B
; FILING DATE: 09-FEB--1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallegos, R. Thomas
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: 75SD1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/952-9881
; TELEFAX: 415/952-2614
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-194-088B-15

Query Match           1.0%; Score 65.6; DB 1; Length 10596;
Best Local Similarity 47.0%; Pred. No. 4,4e-05;
Matches 235; Conservative 0; Mismatches 264; Indels 1; Gaps 1

Qy      2778 cgtacggagagcattccctcagaagccctcgtttctcttcacagcatatgtagaatat 2837
Db      2961 CCTCTCTGACCCCGGGCTTCACCTCTCTGCTCTGCCCCCTCTCTGCTCTGCCCCCTCTCTCT 2902
Qy      2838 ggccgctgcacactgctgcttgccaggactagacacaccccgaaagctgagcttcccg 2897
Db      2901 GGCCGCTGCCCTTCCTGCCCCCTCTGCTCTGCCCCCTTCGCCCCCTTCCTGCTCTGCCCCCT 2842
Qy      2898 acgcttcgtgaacatactatgcttgttggtctctgtggagctggagatcatcccg 2957

```

[illegible]

```

RESULT      10
US-08-194-087-15/c
? Sequence 15, Application US/08194087
? Patent No. 5879910
? GENERAL INFORMATION:
? APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Wela
? TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN
? NUMBER OF SEQUENCES: 21
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Genentech, Inc.
? STREET: 460 Point San Bruno Blvd
? CITY: South San Francisco
? STATE: California
? COUNTRY: USA
? ZIP: 94080
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Path (genentech)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/194,087
? FILING DATE: 18-MAY-1992
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: .
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Dreger, Ginger R.
? REGISTRATION NUMBER: 33,055
? REFERENCE/DOCKET NUMBER: 779
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415/225-3216
? TELEFAX: 415/952-9881
? TELEX: 910/371-7168
? INFORMATION FOR SEQ. ID NO: 15:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 10596 bases
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
US-08-194-087-15

```

[illegible]

```

; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
PCT-US93-04648-15

Query Match      1.0%; Score 65.6; DB 5; Length 10596;
Best Local Similarity 47.0%; Pred. No. 4,4e-05;
Matches 233; Conservative 0; Mismatches 264; Indels 1; Gaps 12.

QY 2778 cgtacgagagcatccctccagcagcctggtcttcttcacatcattgltgaaglat 2837
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2961 cctcctgcagccccggcgctcacctcctgctcctgctgacctcctgcccctcctgcccncct 2902
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2838 ggcgcctgcacctctgcttgaggcagcactagaacacaccgccgaagtgtccocgg 2897
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2901 gcmtcctccccctcctcctccccctcctgctcctgcccctcctgcccctcctgcccctc 2842
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2898 agccttcctgaagaaatactagcttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2957
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2841 cctccccctcctgctcctgctcctgcccctcctcctcctcctcctcctcctcctcctcct 2782
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2958 tgtccacaagaatgtgtgaagtgcgcagacactcaaacgcagcctgtgtttctctcaag 3017
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2781 cctcctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 2722
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3018 tgacctctgtccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3077
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2721 gccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 2662
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3078 gcaagtgtagcagcagcctcagatgcgtccctaalcagcagcctgtgtgaacctgtgcgat 3137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2661 cctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 2602
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3138 ggaattccacgcgatacctgtgcagccaagacactagtgaaaccttaacct-cccctgtct 3196
    || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2601 gccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 2542
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3197 gccctcctgcatacctcagcctgtgcctccctcctctgtgtcctcacaacctcccaagct 3256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2541 gccctcctgcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 2482
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3257 ccacctctcctcagcgaagcc 3276
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2481 cctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 2462
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ. ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZ9PT-F15
US-08-232-463-14
```

```
Query Match 1.0%; Score 61.4; DB 1; Length 7218;
Best Local Similarity 4.2%; Pred. No. 0.00032;
Matches 17; Conservative 229; Mismatches 155; Indels 0; Gaps 0;

QY 2956 cgtgtccacaagaatgtgagctgacgagacacccaagcagcctgcttctcctc 3015
DB 1038 CTGTGCTGCAAGTCAGGAGCTGCGATYYYYYYYYYYYYYYYYYYYYYYYY 1097
QY 3016 agtgacctctgcccctgtgacgagcgttctctcagcctgtgctcgagccactac 3075
DB 1098 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1157
QY 3076 aagcaggtgacgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3135
DB 1158 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1217
QY 3136 atggaattaccgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3195
DB 1218 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1277
QY 3196 tgcacctctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3255
DB 1278 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1337
QY 3256 tccacctctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3315
DB 1338 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1397
QY 3316 ccattcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3356
DB 1398 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTAC 1438
```

```
RESULT 13
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
```

```
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ. ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZ9PT-F15
US-08-232-463-14
```

```
Query Match 0.9%; Score 56.2; DB 1; Length 7218;
Best Local Similarity 5.0%; Pred. No. 0.0045;
Matches 19; Conservative 211; Mismatches 149; Indels 0; Gaps 0;

QY 3946 aaaaatctgatalgaagcgcgctagaggaacacatctgggtaccatcgagct 4005
DB 1411 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1352
QY 4006 cgacaagaatgtgtcggaagtcgtgagagagacccgcttgggaatcagagaactg 4065
DB 1351 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1292
QY 4066 cgcgcgcggaagcgtcacacactggaagcaaacctcagacgcgltgagcaagacaa 4125
DB 1291 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1232
QY 4126 gatgaatggaacagagcgtgtggaaggaacactggcaacagcagcagcagcagc 4185
DB 1231 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1172
QY 4186 gttctgacacactggaatcagcagcagcagcagcagcagcagcagcagcagcagc 4245
DB 1171 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1112
QY 4246 gttctggggcagtgctggaagtcgtgtaagcctggagcagctgccagagtgccctc 4305
DB 1111 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1052
QY 4306 ttctgacagtcgctg 4324
DB 1051 GACCTGACGCAAGCTCGG 1033
```

```

RESULT 14
US-09-046-992-1
? Sequence 1, Application US/09046992
? Patent No. 614006
? GENERAL INFORMATION:
? APPLICANT: Lorberboun-Galski, Haya
? APPLICANT: Yarkoni, Shai
? APPLICANT: Ben-Yehudah, Ahmi
? TITLE OF INVENTION: METHODS OF CANCER DIAGNOSIS
? TITLE OF INVENTION: USING A CHIMERIC TOXIN
? NUMBER OF SEQUENCES: 7
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Pennie & Edmonds, LLP
? STREET: 1155 Avenue of the Americas
? CITY: New York
? STATE: NY
? COUNTRY: USA
? ZIP: 10036-2811
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: Windows
? SOFTWARE: FASTSEQ for Windows Version 2.0b
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/046, 992
? FILING DATE: 24-MAR-1998
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Polissant, Brian M
? REGISTRATION NUMBER: 28,462
? REFERENCE/DOCKET NUMBER: 9457-0013-999
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 650-493-4935
? TELEFAX: 650-493-5556
? TELEX: 66141 PENNIE
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1908 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? FEATURE:
? NAME/KEY: Coding Sequence
? LOCATION: 1...1905
? OTHER INFORMATION:
US-09-046-992-1

Query Match          0.8%; Score 52.8; DB 3; Length 1908;
Best Local Similarity 45.8%; Pred. No. 0.016;
Matches 183; Conservative 0; Mismatches 217; Indels 0; Gaps 0

QY 4706 cggtagaatgataaggaaacaccagaaggaacctgatgtcatcgactcattatagaaa 4765
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 965 CGGCGCGGCTGTGHTGAACCAAGTTCAGACAGTAGTGATCCGCAACGCCCTGGCCAGCCCCG 1024
OY 4766 ttgcccggggtctaccagggtctaccgygcacttcggttgacctggtttgcagacaatggccg 4825
DB 1025 GCAGCGGGCGGCACACTGGGCGGAAGGATCCGAGCACGCCGGAGAAGAGCGCCGTCGTGGCCC 1084
OY 4826 ggaagcacgcgagtcttgcgcaaccacacgcgcgagggccccagtcagtgtgtcgcgcgcccg 4885
DB 1085 TGACCTTGCGCCCGCCCGCAGACGACGACGCTTTGTCGCGGCAAGGACACCGGCAACGACGAGG 1144
OY 4886 ccctcgttgtctagtactcgtccctgtctcagaaagacagcgcacacttgcgcgttgggtctgcg 4945
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1145 CCGGCGGCGGCAACGCGCACGHTGTAGCGTCACTGCGCCGGGTGGCGCGCGGTGAATGCG 1204
OY 4946 ttctcttcagacaatctcatccaacgtgtctagaggagatcgcgcacatcttcagacacatcc 5005
    | |||| | ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

QY	DB	Sequence	Score	DB 2	Length	Gaps
QY	1205	CGGGCCCGGAGACAGCGGCGACCCCTGCTGTGGAGGCGCACTATCCACTGGAGCGGAGT	0.8%	1264	1869	0
QY	5006	tgtcgccgaagagagggcttcgtctccgggaagacacttaactgaactggtggtgtag	45.5%	5065		
DB	1265	TCTCGGCGGAGCGGCGGCGACGTCACCTTCAGCACCCGCGGCGGACGAACGAGCGGTG		1324		
QY	5066	ggttgctggaacagcagccggtacttcaactggtggcg		5105		
DB	1325	AGCGGCTGCTCCAGCGCACCGCCAACTGAGAGAGCGCGG		1364		
<p>RESULT 15</p> <p>US-08-356-786-15</p> <p>Sequence 15, Application US/08356786</p> <p>Patent No. 5877305</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Huston, James S.</p> <p>APPLICANT: Opperman, Hermann</p> <p>APPLICANT: Huston, L. L.</p> <p>APPLICANT: Ring, David B.</p> <p>TITLE OF INVENTION: Biosynthetic Binding Protein for Cancer</p> <p>TITLE OF INVENTION: Marker</p> <p>NUMBER OF SEQUENCES: 16</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: Edmund R. Pitcher, Testa, Hurwitz, &amp; Thibault</p> <p>STREET: Exchange Place, 53 State Street</p> <p>CITY: Boston</p> <p>STATE: Massachusetts</p> <p>COUNTRY: USA</p> <p>ZIP: 02109</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Floppy disk</p> <p>COMPUTER: IBM PC compatible</p> <p>OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>SOFTWARE: Patentin Release #1.0, Version #1.25</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/08/356,786</p> <p>FILING DATE:</p> <p>CLASSIFICATION: 424</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: 07/831,967</p> <p>FILING DATE: 06-FEB-1992</p> <p>ATTORNEY/AGENT INFORMATION:</p> <p>NAME: Pitcher, Edmund R.</p> <p>REGISTRATION NUMBER: 27,829</p> <p>REFERENCE/DOCKET NUMBER: CRP-053</p> <p>TELECOMMUNICATION INFORMATION:</p> <p>TELEPHONE: (617) 248-7000</p> <p>TELEFAX: (617) 248-7100</p> <p>INFORMATION FOR SEQ ID NO: 15:</p> <p>SEQUENCE CHARACTERISTICS:</p> <p>LENGTH: 1869 base pairs</p> <p>TYPE: nucleic acid</p> <p>STRANDEDNESS: single</p> <p>TOPOLOGY: linear</p> <p>MOLECULE TYPE: DNA (genomic)</p> <p>FEATURE:</p> <p>NAME/KEY: CDS</p> <p>LOCATION: 1..1869</p> <p>OTHER INFORMATION: /note= "product = "741SFV-PE40""</p> <p>US-08-356-786-15</p>						
QY	4706	cggtagaagatgaagaaacaccagagagaccctgagatgctatcagacatcatgatacaga	0.8%	4765		
DB	926	CGGCGCGGCTGTCGTGGAAACGAGTCAACAGAGTGATCGCAAGCGCCCTGGCAGCGCGG		985		
QY	4766	ttgccgggggtaccagggtctaccggacattcggtcgtgaactgtgttcagaacatggcgg		4825		

```
Db 986 GCAGCGCGCGGACCTGGGGGAAGCATCCGCGAGCAGCCGGAGCAAGGCCGCTTGGCCC 1045
QY 4826 ggaagacagcgagagctggggaaccacgcccagtcgcatgtygcaagcgccg 4885
    || || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1046 TGAACCTGGGCCGCCCGAGAGGAGCGCTTCTCCGCAAGGCAACCGGCAAGAGAGG 1105
QY 4886 ccctcgtgctgagtaacctcgccctcgcgagagaccagcgcaacctgcccgtgagctgcg 4945
    || || || | | | | | | | | | | | | | | | | | | | | | | | |
Db 1106 CCGGCGCGGCGCAACGCCGACGTGTGAGCCTGACCTGCCCGGTCGCCCGGCGGAATGCG 1165
QY 4946 ttcccttcagaaacatcatcaccagtytctagagagatcgcccatctcgaagacatcc 5005
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1166 CGGGCCCCGGCGGACAGCGCGGACGCCCTGCTGGAGCGCAACTATCCCACTGGCGCGAGT 1225
QY 5006 tgcgcgcgacgagagagagcttcctgcgcggaagcacttcactgagctgggctgtag 5065
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1226 TCCTCGGCGGACGCGGCGAGGTCAAGCTTCAGCAACCGCGGCAACGCAAGACTGGACGCTGG 1285
QY 5066 ggttgctggaacagcgagccgctacttcaacatgggcgg 5105
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1286 AGCGGCTGCTCCAGGCGCACCGCCCAACTGAGAGGCGCGG 1325
```

Search completed: October 5, 2002, 07:37:58  
Job time: 18460 sec